Universidad Autónoma de Madrid

Facultad de Medicina
Departamento de Bioquímica

Novel algorithms adding new dimensions to mass-spectrometry based proteomics: comprehensive characterization of post-translational modifications.

Doctoral dissertation

Navratan Bagwan

M.Sc. in life science informatics

Thesis coordinators

Prof. Jesús Vázquez Cobos
Dr. Elena Bonzón-Kulichenko

Centro Nacional de Investigaciones Cardiovasculares (CNIC)

Madrid, 2018

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union´s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement nº 608027 (¨CardioNext¨ Initial Training Networks project)

Acknowledgment

...Until we meet again

Those special memories of you
will always bring a smile
if only I could have you back
for just a little while
Then we could sit and talk again
just like we used to do
you always meant so very much
and always will do too
The fact that you're no longer here
will always cause me pain
but you're forever in my heart
until we meet again

I would like to dedicate my thesis to
the memory of my father…

I want to thank many people, without their invaluable academic, educational, psychological and human support and belief in me as a researcher; this work would not have the spirit that it has. Firstly, I would like to thank my supervisor Jesús Mª Vázquez Cobos, who has given me the opportunity to do this PhD. I consider myself fortunate enough that I got an opportunity to work under world-leading proteomic researcher.

Thank you Dr. Jesus, my deep gratitude goes to you. You always guided me throughout this journey and shared the excitement of long years of discovery. Your unwavering enthusiasm for science kept me constantly engaged with my research. I am especially thankful to you for having the faith in my research, and me. I really appreciate the motivation you provided during the tough times of my PhD and personal life. Most importantly, other than science, I will always be inspired by your enthusiasm about everything. And yes, one day will and we should definitely plan a trip to India.

I would like to thank my co-supervisor Dr. Elena Bonzón Kulichenko, for the immense support throughout this time. You have taught me that with right attitude, patience and hard work towards our research, one can eventually seek success. Your enthusiasm, commitment and active contribution to research, has always inspired and motivated me, especially during the rough times. While working under your guidance I have evolved both at personal and professional level. I will always value the things you taught me, let it be small or big. Now when I have my work in my hands ready and reaching to the final step, looking back in times, I can confidently say this: without you, this would not have been possible. You have been a mentor to me in true meanings.

Moving ahead, I am extremely grateful to my friends and colleagues in “la cueva” for providing love, support and a sense of being and belonging there, which often saved me from burnout. Thanks everyone: Marco for all the help, chitchat and humour we shared and of course, the stories of your childhood in India made me amazed and happy. You are the first one in Spain who made me feel home and provided all the guidance and support I needed. Iakes, thanks mate, for everything; I learnt things outside science, which I would have never learnt anywhere else. Spiros, it was probably the best time in cave when you were around, thanks for being such a great friend I still think of somethings and wonder what Spiros would say. Enrique, thanks for many things: I have to say you never said no to me for anything, you thought me, explained me things. Despite all the ups and downs, I am sure I made a good friend (now finally it’s time for you to cut my name from your list). Juan Antonio, “que bien vives”, thanks for always

being supportive, funny, and nice. Emilio, the journey of PTMs and FDR actually started with you and I remember the way you helped and explained, i am glad that I got to work with you, thanks for everything. A big thanks to all the people in the lab for always being nice and supportive, I will admire this.

I also would like to thank to my Tutor Dr. Belen Peral for all the help throughout the university process and bureaucracy. Thanks to my collaborator in Germany, Dr. Dominik and Dr. Adam, It was short but a very productive time. Thanks for José Antonio Enríquez for providing the data for this work and a great collaboration. Especial thanks to Lilit and Eduardo, without you people I would have been most likely an illegal immigrant, and never survived the process.

I also want to thanks Surya Gupta; I have come a long way in this journey and of course with loads of sidetracks and always found you there to talk and for support. So thanks for all the support, understanding, help and love.

Last but not the least, my deep and sincere gratitude to my family for their continuous and unappalled love, help and support. I am grateful to my mother for always being there as friend and mentor in life. Thanks “Bhai” for all the love and care you have for me. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. They selflessly encouraged me to explore new directions in life and seek my own destiny. This journey would not have been possible if not for them, and I dedicate this milestone to them.

Abstract

The technological advances in proteomics are allowing an increasingly detailed characterization of the complex panorama of post-translational modifications of proteins and are gradually developing towards an unbiased analysis of peptide modifications. The recently developed ultra-tolerant database search (open search, “OS”) uses precursor mass tolerances of hundreds of Daltons, allowing the identification of modified peptides never identified before by conventional (closed, “CS”) searches. Despite these improvements, OS algorithms still rely on the chance that the modification leaves enough unaffected fragment ions, thus identifies only half of the modified peptides and cannot pinpoint the modification site. Furthermore, there is a need of a generic quantification algorithm able to handle the huge variety of modified peptides resulting from an OS experiment.

In this Thesis, I present a suit of developed algorithms and tools, designed to overcome the above-mentioned limitations. Comet-PTM is an improved search engine that applies the peptide modification mass to the fragmentation series upon score calculations for each peptide-spectrum match (PSM). As a result, we emulate the scores produced by a CS for the same modification set as variable; double the yield attained by a regular OS and localize the modified residue with high accuracy. SHIFTS, controls the PSM false-discovery rate of the CometPTM results through a conservative three-layered approach taking into account the high mass accuracy of modern mass spectrometers.

PtmSticker annotates the enormous wealth of modifications in a semi-supervised way, allowing for the first time the generation of a complete map of the modified peptidome as part of an automated pipeline. For the quantitative analysis of modified peptides, we developed and validated an algorithm based on a previously proposed WSPP workflow, for the simultaneous quantification of the modified peptidome, the whole proteome and systems biology. The model allows detection of PTMs changing independently of the protein abundance change.

These developments were used to characterize the impact of mitochondrial heteroplasmy on the proteome and on the modified peptidome in mice, revealing that the heteroplasmy causes oxidative damage in heart OXPHOS proteins.

Resumen

Los avances tecnológicos en proteómica permiten una caracterización cada vez más detallada del complejo panorama de las modificaciones postraduccionales de proteínas y están derivando hacia un análisis no sesgado de las mismas. Los recientes algoritmos de búsquedas ultra-tolerantes en bases de datos, denominados búsquedas “abiertas, emplean tolerancias de cientos de Dalton alrededor del precursor y permiten la identificación de péptidos modificados que nunca antes se habían identificado en búsquedas convencionales. A pesar de estos indudables beneficios, para identificar un péptido modificado, las búsquedas abiertas siguen necesitando que la modificación deje suficientes fragmentos no-modificados. Por esta razón, sólo se identifica la mitad de los péptidos modificados en comparación con una búsqueda convencional, y no se llega a identificar el sitio de la modificación. Además, hasta la fecha no existe ningún modelo estadístico apropiado para la cuantificación de la enorme variedad de péptidos modificados que resulta de una búsqueda abierta. En esta tesis se presenta un conjunto de algoritmos y herramientas informáticas que resuelven todas estas limitaciones, llevando las búsquedas abiertas a un nivel superior de desarrollo. El Comet-PTM es un buscador en bases de datos capaz de tener en cuenta la masa de la modificación para aplicarla a los fragmentos a la hora de calcular la puntuación del par espectro-péptido (PSM). Como resultado, se logra emular la puntuación que se obtendría en una búsqueda convencional con esa misma modificación definida como variable. Además, se duplica rendimiento de PSMs de la búsqueda abierta y se localiza el sitio de la modificación. SHIFTS es una herramienta para controlar la Tasa de Falsos Descubrimientos, FDR *(False Discovery Rate*) de los resultados del Comet-PTM, mediante una aproximación conservativa de 3 capas que tiene en cuenta la alta exactitud de masa que caracteriza a los equipos de espectrometría modernos. El PtmSticker anota las modificaciones de forma semi-supervisada, permitiendo por primera vez obtener un mapa completo del peptidoma modificado como parte de un flujo de trabajo automático. Sobre la base de un modelo estadístico existente en nuestro laboratorio, el WSPP, hemos desarrollado y validado un algoritmo para la cuantificación del “peptidoma modificado” de forma simultánea con el proteoma y el análisis de biología de sistemas. Finalmente, la plataforma Comet-PTM desarrollada en esta tesis ha sido aplicada al estudio del impacto de la heteroplasmia mitocondrial sobre el peptidoma modificado en diferentes tejidos de ratón. Nuestro estudio ha revelado que la heteroplasmia produce un daño oxidativo a las proteínas de la cadena de transporte mitocondrial fundamentalmente en el corazón.

Table of Content

Abbreviations

term	description
AGE	Advacance Glycation Products
ANNOVA	Analysis of Variances
CS	Close Search
CVD	Cardiovascular disease
ESI	Electrspray ionization
FASP	Filter Aided Sample Prepration
FDR	False Discovery Rate
GIA	Generic Integration Algorithm
HILIC	Hydrophilic Interaction Chromatography
HPLC	High Performance Liquid Chromatography
IAM	Iodoacetamide
iTRAQ	isobaric Tags for Relative and Absolute Quantitation
m/z	Mass-to-charge
MAD	Mean Absolute Deviation
MS	Mass Spectrometery
MS/MS	tandem mass spectrometry
MS1	survey scan (or full scan)
MS2	fragmentation spectrum
OS	Open Search
OXPHOS	Oxidative Phosphorylation System
PPM	Parts Per Million
PSM	Peptide Spectrum Match
PTM	Post Translational Modfication
SBT	Systems Biology Triangle
SIL	Stable Isotop labelling
TMT	Tandem Mass Tags
WSPP	Weighted Spectrum, Peptide, Protein [statistical model]

Introduction

1. What are post-translational modification (PTMs) of proteins?

Proteins are fundamental biomolecules or macromolecules of the cell that carry out and regulate major biological functions. More importantly, they help to maintain the structure/shape of the cell and work as signal transmitters between inner and extra cellular environment. PTMs of proteins can be defined as *in vivo* biochemical mechanism in which the amino acid residues are covalently modified after protein translation (Prabakaran *et al.*, 2012). PTMs can take place at different amino acids and the process is often driven by an enzymatic activity/reaction. In fact around 5% of the proteome is known to have enzymes that can result in hundreds of PTMs, and many of these PTMs are found to be linked with human disease-associated mutations (Duan and Walther, 2015). Studies of PTMs ranging from yeast to human represents how important they are and known to control or modulate the functional activity of proteins (Mann and Jensen, 2003; Oliveira and Sauer, 2012). For instance, approx. half of the proteins in *Saccharomyces cerevisiae* are known to be affected by phosphorylation in a metabolic network (Chi *et al.*, 2007). On the other hand, in humans, abnormal phosphorylation has been studied in many diseases, such as cancer and Alzheimer (Khadjavi *et al.*, 2011).

Unlike the changes in gene sequence which occurs on an evolutionary time scale and is not required for instant physiological change or development, PTMs permit amino acid properties to be changed “on the fly” if needed for specific response such as the continual battle against the disease (Prabakaran *et al.*, 2012).

PTMs on the different amino acid side chains or at protein C- or N- terminal can extend chemical arrangement up to 20 amino acids by modifying or adding different functional groups such as phosphate. The most studied functional groups added-on to amino acids are oxidation, acetylation, phosphorylation and ubiquitination. In many cases, multiple amino acids are modified by multiple PTMs, which leads to a combinatorial explosion of possible molecular and chemical states. Possible combinations of such PTMs are often the epicentre of sophisticated forms of cellular information processing that are essential for the emergence of organismal complexity.

2. Importance of PTMs in biological systems

The current advances in the field of systems biology and proteomics, has opened a new era of research where much of the focus has been on PTMs. In recent years the number of published papers has been increasing drastically (Figure 1).(Olsen and Mann, 2013).

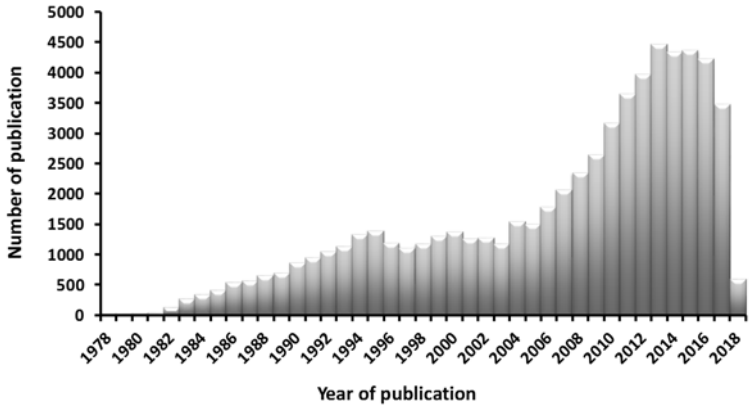


Figure 1: Bar plot, representing the number of papers published related to PTMs (PUBMED, keyword; "Post Translational Modification") in the last three decades.

The growing need to understand and predict the molecular cause of disease remain as a major bottleneck of current medicinal biology. However, taking into account the importance of PTMs, a continuous interest is developing towards computational analysis of disease associated amino acids modification. To this end, various studies have successfully demonstrated the role of PTMs in different diseases such as cardiovascular disease (CVD), aging, mitochondrial dysfunction and Alzheimer.

CVD is one of the leading causes of mortality and morbidity all around the world (Roth *et al.*, 2015). There has been a great emphasis on PTM (Carbamylation, Glycation and Oxidation, phosphorylation) research as they are involved in pathogenesis of CVD (Santos and Lindner, 2017) and eventually lead to progression of other forms of chronic disease such kidney disease (Sadayappan *et al.*, 2009; Gupta and Robbins, 2014; Gajjala *et al.*, 2015).

In other studies, acetylation has been linked to number of diseases, for example, in cancer where N-acetyl provides a unusual subcellular distribution of the tumour antigen (Chammas *et al.*, 1999). In dementia, acetylation impairs synaptic plasticity (Mattson, 2010; Pirooznia and Elefant, 2013). Glycosylation is another well-studied PTM, which results from the addition of sugar to a protein. A class of glycosylation, O-linked glycosylation has been proven to have a positive effect on Alzheimer's disease, as it reduces the neurofibrillary tangles in neurons

(Broncel *et al.*, 2010). Advanced Glycation End products (AGEs) are also targets in cardiovascular pathologies, and are commonly used a biomarkers, since AGEs have a key role in the progression of cardiovascular diseases in patients with diabetes (Hegab *et al.*, 2012).

Moreover, novel medical technologies aiming to prevent mitochondrial genetic diseases, to improve fertility, or to deliver foreign mitochondria with therapeutic goals, can produce mitochondrial heteroplasmy. Mitochondrial heteroplasmy is the presence of more than one variant of mtDNA in the same cytoplasm. In collaboration with the lab of José Antonio Enríquez Domínguez, efforts are made to facilitate a mechanistic model (mouse model), explaining the role of PTMs and understand why the heteroplasmy is harmful in tissues that do not resolve it.

Without any doubt, enormous efforts are made to get a complete glimpse of PTMs in biological systems, but we are still far away from understanding and unravelling the combinatorial complexity of PTMs. Technologies such as mass spectrometry (MS) -based proteomics, are definitely increasing the quality of PTM identification and also allowing us to identify new modifications of proteins. With the constant increasing list of new modifications, the dynamics of protein function needs to be re-evaluated, as most studies concerning PTMs start with a prior knowledge or hypothesis, focusing only on a limited number of PTMs, hence neglecting the completeness of proteome. If the personalized medicine or protein-based biomarkers are the next big platform, it is important to develop tools and methods, which can identify all PTMs in a given biological system, thereby facilitating the complete study of the nexus between the proteome and PTMs.

3. Second generation (Shotgun) MS based proteomics

The term shotgun proteomics refers to the bottom-up proteomics approach where peptides from digested protein are identified using a combination of a high performance liquid chromatography (HPLC) and mass spectrometry. In a typical workflow (Figure 2), proteins are extracted and denatured either in urea or by high temperature incubation with sodium dodecyl sulphate (SDS). Disulfide bonds are reduced, for example with dithiothreitol (DTT), and alkylated, usually with iodoacetamide (IAM), to prevent protein renaturation. Afterwards, proteins are digested, typically with trypsin (Lina P. Aristoteli, Mark P. Molloy, † and Mark S. Baker*, †, 2006), and the resulting peptide mixtures are fractionated by different means, such as high pH reversed-phase or strong cation exchange chromatography (Junmin Peng *et al.*, 2002).

Peptides are further separated by reversed-phase high-performance liquid chromatography (HPLC) on-line with MS. As peptides elute from the chromatographic column, they are ionized by electrospray ionization (ESI) and injected into the MS instrument. The mass-to-charge (m/z) ratios of the peptide ions are measured to determine the molecular mass of each peptide, producing the MS spectrum. Then, from the top ten most abundant ions, each peptide ion is isolated in the first mass analyzer (MS 1) and directed into a collision cell, where it collides with neutral gas molecules (for example, helium) and becomes fragmented. The mass-to-charge ratios of the resulting fragments are measured in the second mass analyzer (MS 2), producing a tandem mass spectrum (MS/MS spectrum). Fragmentation of ions is mostly performed by collision induced dissociation (CID) (Mitchell Wells and McLuckey, 2005) or higher energy dissociation (HCD) (Ferries *et al.*, 2017). In these fragmentation techniques, peptide precursor ions are broken into pieces at their peptide bonds, and the resulting fragment ions have mass differences corresponding to the residue masses of the respective amino acids. To identify the resulting fragments, the Roepstorff-Fohlman annotation is generally used, where, ions from the N-terminus of the peptide are known as *b* ions, while fragments from the C-terminus are called *y* ions (Roepstorff and Fohlman, 1984). This cycle of alternating MS and MS/MS spectra is repeated throughout the chromatographic run, for what is known as the data-dependent acquisition (DDA).

Following the data collection, the MS/MS spectra are further used for high throughput identification of peptide and proteins in a database using different computational algorithm such as Sequest (Eng, McCormack and Yates, 1994), comet (direct descended of academic version of Sequest) (Eng, Jahan and Hoopmann, 2013a), Mascot (Perkins *et al.*, 1999) and X!!Tandem (Bjornson *et al.*, 2008). In the database search engines, each experimental spectrum is matched and scored against the most plausible theoretical peptide in the database (Figure 2).

3.1. Conventional database search approach for peptide identification

A crucial component of the analysis of shotgun proteomics datasets is the search engine, an algorithm that attempts to identify the peptide sequence from the parent molecular ion that produced a certain fragment ion spectrum in the dataset. For simpler understanding of this process, database search can be categorized in three basic steps: 1) collection of all MS/MS spectra and querying a sequence database for peptides (theoretical peptide) of specific mass; 2) scoring the theoretical peptide against experimental peptide; 3) providing validation for the putative peptide identification. Briefly, for every MS/MS spectra, experimental peptide mass is calculated using its *m/z* and charge. The database search

algorithm then collects the set of theoretical peptides having similar or approximate masses to the experimental peptide. For every candidate peptide, fragment ion masses are calculated, matched with the ones of the experimental spectrum, and the peptide candidates are ranked according to their peptide spectrum match (PSM) score.

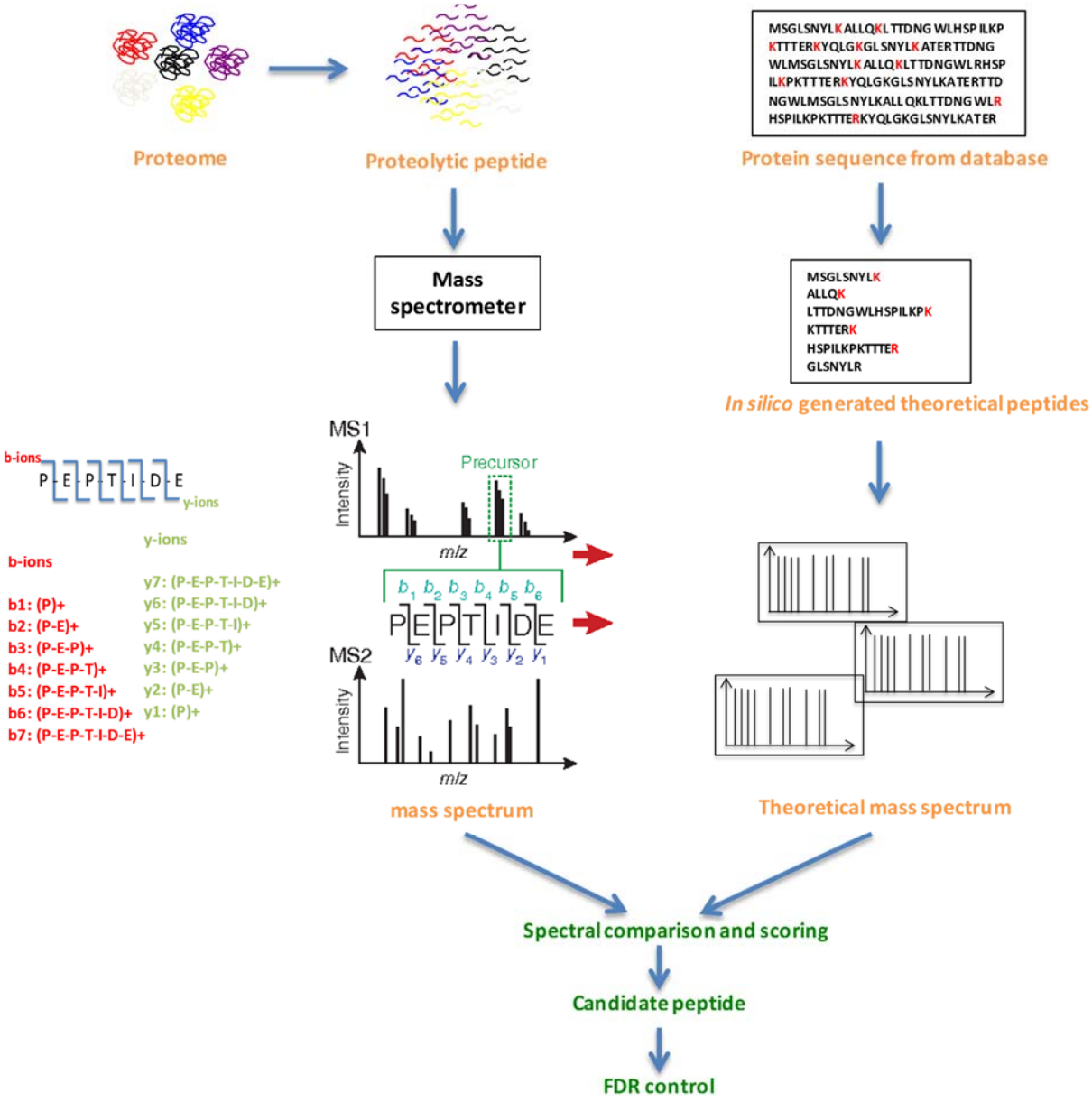


Figure 2: schematic representation of a typical shotgun proteomics experiment. Left side showing the process where tryptic peptides are generated from a biological protein sample. The obtained peptide mixture is usually fractionated/or not and sent to mass spectrometer usually through a coupled RP-HPLC. The resulting data mainly provides MS1 and MS2 spectra, where MS1 represents intensity versus the m/z of the ionized components at a given retention time, and MS2 contains the intensity versus the m/z of the fragmented products of a component from MS1 at given retention time. The produced data is then searched against theoretical mass spectra of peptides generated from in-silico digested proteins from a database of choice. Upon matching the theoretical and experimental spectra, every match is scored and results in a candidate peptide. Then, using the target/decoy approach, candidate peptides are filtered out by false discovery rate.

3.1.1. Identification quality scores

For ensuring the quality of identification from MS based experiments, it is necessary that the search engine has a scoring function that consistently ranks correct PSMs better than the poor matches. In the past, there have been multiple publications and tools suggesting different strategies. One of the most widely used approaches is from Sequest (Eng, McCormack and Yates, 1994), where scores are determined on how well the experimental and theoretical spectra are matched/overlapped using cross correlation, which is a sum of all intensities between theoretical and experimental peptide. Being robust and fast, Xcorr function form Sequest is being widely used and studied in proteomics. The first version of Sequest Xcorr was calculated in a two-step process, where, initially a preliminary score was calculated and followed by a cross correlation for the matches which passed a preliminary score.

Sequest Xcorr has experienced significant improvements over the years (Eng *et al.*, 2008; Klammer, Park and Noble, 2009; Diament and Noble, 2011). Particularly, a new search engine, Comet (Eng, Jahan and Hoopmann, 2013b), has implemented the Sequest algorithm with an improved cross correlation (Eng *et al.*, 2008). Comet avoids using Fast Fourier Transformation (FFT) to calculate the full cross correlation, which is computationally expensive. Instead, a pre filter of a preliminary score is used. In addition, the Xcorr calculation is simplified and is calculated for every candidate peptide, associating an *E*-value for the significant match. *E*-value increases the sensitivity and allows a simple comparison among scores of spectra, where *E*-value represents the number of peptides that are expected to match and score well above the best match just by chance.

On the other hand, some approaches use statistical probability methods to compute quality of match, such as MASCOT (Perkins *et al.*, 1999). In MASCOT, the score is the probability that the observed match between the experimental and theoretical peptide is just by chance. The open source search engine Andromeda (Cox *et al.*, 2011) follows an approach similar to MASCOT.

3.2. PTM identification: limitations of conventional search engines

More than 500 PTMs have been described on UniProt database (April 2018). Hundreds of them have been described in *Homo sapiens.* Over the years, PTM identification study was mainly limited to a protein or a specific pathway since most of those studies started with a

specific question and a hypothesis. Technological advancements in MS has encouraged scientist to perform database searches with increasingly narrower precursor and fragment mass tolerance, which is usually set in the parts-per-million (PPM) range (Bonzon-Kulichenko *et al.*, 2015). The need to discard incorrect matches and faster database search are one of the few reasons behind the use of narrow tolerance (Hsieh *et al.*, 2010; Bonzon-Kulichenko *et al.*, 2015). However, considering the fact that the number of possible peptide modifications is overwhelming and impossible to anticipate, the database searches aiming to identify modified peptides possesses a greater challenge if narrow tolerances on precursor and fragments are used. In most cases, common chemical modifications, such as methionine oxidation or cysteine alkylation, are added to the search, together with some known biological PTMs. However, the mass shifts produced by unknown modifications are not considered during the search, resulting in lower PSM scores and/or miss-assignments.

3.3. Unwanted sources of peptide modification

For a number of reasons, MS spectra generated from peptides with modifications outside of the user-specified rules may be present in tandem MS datasets. Non-protease-specific peptides could arise from in-source fragmentation (Kim *et al.*, 2013); chemical artifacts introduced during sample preparation (Savitski, Nielsen and Zubarev, 2006), for instance, over-alkylation (and and Fales*, 2001). Studies have shown that tryptophan oxidation may occur following protein purification and isolation, particularly with the use of gel electrophoresis. Moreover, electrospray ionization has also been reported to induce oxidation of methionine, tryptophan or tyrosine residues (Potgieter *et al.*, 1997). In addition, the labile nature of some modifications also makes them difficult to identify by MS. For instance, oxidative modifications, glycosylation and phosphorylation (on serine and threonine residues) are labile modifications and modified peptides can undergo loss of the substituent under conditions used in electrospray ionization (Carapito *et al.*, 2009; Chick *et al.*, 2015).

In addition, peptides having more than one residue capable of bearing a modification, for instance, peptide with phosphorylation with multiple serine and threonine residues. In such cases, during the spectrum match, the confidence of localizing a modification to a particular residue is dependent on the presence of fragment ions in the MS/MS spectrum derived from the fragmentation between the possible sites of modification. Conventional search engines currently pay little or no attention to representing the certainty or ambiguity of modification

localization in their top scoring result. Considering all the limitations in the existing algorithms mentioned above, only a small fraction of modified peptides is identified and only in a targeted way. Thus, there is a need of an algorithm that would anticipate the mass shift present in the precursor and fragments.

3.4.Hypothesis-free modified peptides identification and the dark matter

In the age of MS-based high-throughput proteomics (Link *et al.*, 1999), despite having access to millions of spectra, most experiments start with a hypothesis driven approach, especially considering analysis of PTMs. In such cases, the studies tend to consider a subset of modification and the unmodified form of the peptide to draw a conclusion and leaves out the possibilities of having a complete picture of a biological system. Hence, despite the advances in speed and sensitivity which allow the generation of millions of spectra per experiment, only a minority of these spectra could be mapped to proteins (Skinner and Kelleher, 2015; Griss *et al.*, 2016). A large proportion of unassigned spectra are thought to arise from peptides containing sequence variants or unknown chemical and posttranslational modifications (PTM) (Griss *et al.*, 2016), and their characterization is one of the most interesting and challenging goals in proteomics. A number of computational methods have been proposed for the detection of these unmatched peptides (Marshall Bern, Yuhan Cai and Goldberg, 2007; Chen *et al.*, 2009; Kim and Pevzner, 2014; Ma and Lam, 2014; Shortreed *et al.*, 2015; Griss *et al.*, 2016).

Recently, an “open search” (OS) strategy, where precursor mass tolerances of hundreds of Da were used with a conventional search engine, was reported to identify modified peptides at an unprecedented scale (Skinner and Kelleher, 2015). OS established that many of the unexplained/unidentified spectra could be explained: by unexpected PTMs or rarely but possibly, from amino acid substitutions not present the search database. They used an ultra-wide tolerant of <1000 Da in Sequest search that identifies hundreds of different type of chemical and biological PTMs.

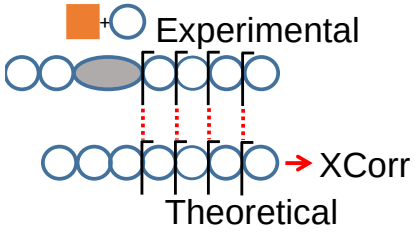


Figure 3: : Conventional OS methods can identify modified peptides from MS/MS spectra, but only the fragments unaffected by the modification (Orange Square) are matched; this effect diminishes the score assigned to modified peptides, decreasing identification performance

Another report demonstrated that OS could be performed at orders-of-magnitude faster speeds using a novel fragmentation-ion indexing algorithm (MSFragger) (Kong *et al.*, 2017). MSFragger uses the indexing of nonredundant set of peptide and then with mass binning and mass ordering in fragment index to facilitate the faster retrieval of candidate peptides. MSFragger demonstrated it use for OS even with variable modification.These two methods may have a considerable impact on the field, opening the way to true hypothesis-free analysis of PTMs by MS; however, OS algorithms still rely on the chance that the modification leaves enough unaffected fragment ions for matching by the search engine (Figure 3). OS strategies can therefore identify only approximately half the modified peptides detectable by conventional "closed" searches (CS) (Chick *et al.*, 2015). Moreover, existing OS approaches cannot directly identify the modification site.

4. Identification confidence

Conventional search engines in the era of high throughout proteomics enabled the identification of hundreds of thousands of peptides in MS/MS data on a routine bases, but at the same time made validation of identifications a tedious process. Thus, the false discovery rate (FDR) is generally used to control the percentage of the possible wrong identifications. FDR estimates the number of false positives in a given dataset, which is the ratio of the number of incorrect PSMs and in the total number of PSMs at a chosen identification score threshold (Choi and Nesvizhskii, 2008; Nesvizhskii, 2010).

4.1. Target/Decoy approach

Most accepted and widely used strategy to evaluate the FDR is based on the target/decoy approach (Elias and Gygi, 2007; Bonzon-Kulichenko *et al.*, 2015), where spectra a searched nor only against the target data base, but also against a "decoy" or false database

of the same size as the target data base. The decoy database is made *in silico* from the target database model organism by either reversing or by shuffling the tryptic peptides. The idea behind this approach is that the PSMs that hit the decoy database are false, and the number of these false PSMs is equal to the number of false PSMs from the target database at a certain score threshold. Creating a decoy is one of the most important factors, which can directly create a bias in identification sensitivity. Moreover, creating a perfect decoy database is almost impossible without losing the composition of target sequence. For these reasons, decoy database made by shuffling or reversing the target sequence, should maintain a target sequence composition, for instance, decoy database should have a similar amino acid distribution and length distribution as target protein database. Moreover, there should only be one decoy sequence for every target counterpart, as this helps to reduce the chances of randomisation and will help to get reproducible results (Elias and Gygi, 2007; Jeong, Kim and Bandeira, 2012).

These target/decoy databases can be used in a concatenated or separated form. In the separated search, both target and decoy databases will be searched separately, so that the score distribution form both searches can be assessed and used for FDR calculation. In the same stream, a concatenated search is conducted by combining the decoy and target databases together. This search reportedly has many benefits, starting with a reduced search space around 20% and an increased sensitivity, since it gives a fair chance of competition between target and decoy match, and only the best match will be chosen for each scan from either target or decoy database(Elias and Gygi, 2007).

4.2. Concerns in PTM FDR

One of the major concerns regarding PTM identification and FDR control is that the FDR for modifications can drastically vary from the unmodified form of peptides. Some recent studies (Fu, 2012; Fu and Qian, 2014; Kong *et al.*, 2017; Li *et al.*, 2017) suggested solutions and the discussed possible reasons for such cases. In most proteomics studies only a subset of PTMs are defined as variable modification and a global FDR is calculated and raises the risk of under or over estimating FDR. The reasons of happening are that: 1) spectra of modified peptides can have their own features (e.g. insufficient fragmentation or neutral losses), thus, can have different score distributions from those of unmodified peptides, 2) the proportions of modified and unmodified peptide are different in a sample, so the chances of obtaining a correct peptide sequence are different. In a similar way, as the proportions are different for candidate peptides in the search space, the prior chances of

having an incorrect identification are also different. Therefore, a separate FDR also known as "local FDR" has been suggested and used in recent studies.

Alongside the "local" FDR issue, using multiple variable modifications in the search can also increase the search space drastically, which results in two main problems. On one hand, in a huge time complexity. For example, searching for phosphorylation on serine and threonine and oxidation on methionine in peptide "YFDSTDYNMAK" would generate 2^5=32 possibilities. Considering this same scenario for thousands of peptides would lead to search space orders of magnitude bigger than a search without PTM. On the other hand, because of an increased search space, the FDR increases as the probability of matching a false identification just by chance increases. Another parameter in the search engine that can drastically influence the quality of identification is the precursor mass tolerance; for instance, in ultra-tolerant database search, specifying it to a higher error rate will allow a bigger search space, which will result in reduced sensitivity as chances of random matching increases.

5. Quantification of peptides and proteins

The simplest definition of quantitative proteomics can be defined as the "study of overall or global changes in protein expression". However, the nature of MS based proteomics is not inherently quantitative; a large number of variables contribute to the peak intensity in a mass spectrum such as: absolute amount of protein, relative amount of protein among samples. In addition, proteins with similar concentration may generate completely different spectra and signal intensities. For instance, peptides generated from a protein digestion may produce different signals or some peptides might not be identified. One of the most tedious bottlenecks of current high throughput proteomics, where a large number of differential expression data is collected, is to determine and discriminate whether the observed change in protein is statistically significant or it has been stacked up from PSM or peptide level outliers.

Despite all the hurdles, MS can be used for generating quantitative data. In brief, quantification can be classified in two major categories: stable isotope labelling (SIL) and label free (LF). In the most general way, in the LF technique, for each sample a separate analysis is performed comparing peptide peak intensities. Wherein each protein is then quantified by, taking the most intense peptides averaged from three different technical replicates, or by using spectral or peptide counting, by taking the number of spectra or peptides for each protein.

Most importantly, in any of the label free techniques the reproducibility of the reverse-phase HPLC-MS instrument is critical. There are several statistical models previously published taking into account all the limitations of such a technique, some of the most prominent models uses an analysis of variance (ANOVA) (Daly *et al.*, 2008; Clough *et al.*, 2009; Karpievitch *et al.*, 2009). In brief, label-free quantitation involves analyzing several technical replicates of the same sample or biological replicates belonging to more than one different conditions. The need of having several replicates helps to overcome the difference in ionization efficiency that arises from the same peptide from the different samples analyzed separately in the MS. Because of label free data's multiplicative nature, different factors such as fixed and random error source come into play, thus the quantitative data is usually subjected to logarithmic transformation. Such transformation allows treating the error source as additive, which makes it easier to model the replicate structure within the analysis of ANOVA.

On the other hand, stable isotope labelling (SIL) includes three major categories: metabolic labelling such as SILAC which refers to isotope labelling of amino acids in cell culture (Ong and Mann, 2007). Enzymatic labelling such as ^{18}O (Xudong Yao *et al.*, 2001) and compares only two samples per experiment. Chemical labelling such as iTRAQ (Ross *et al.*, 2004) or TMT (Thompson *et al.*, 2003) are one of the most widely used techniques for large comparative experiments. In SIL, since the same peptide from the different samples ionizes at the same time, it has the same ionization efficiency. In comparison to label free, SIL provides improved quantitative precision, accuracy.

In general there are many statistical models to study and to deal with the variation and outliers in SIL approaches (Chee Sian Gan *et al.*, 2007; Hultin-Rosenberg *et al.*, 2013). Most of these studies use a simplistic way of calculating the protein values as an average (mean or median) of peptide ratios and comparing them among different experiments or replicates using a student's *t*-test or advancing slightly by ANOVA. Moreover, most models and statistical significance tests are based on normality assumptions that have not been tested despite the fact that heterogeneity of variance has been documented in all SIL methods. These techniques are based on peptide-centric measurements, and the lack of general models leads to the subjective choice of a method for combining multiple peptide readings to estimate protein ratios. For example, ANOVA, which considers the same number of scans per peptide and the same number of peptides per protein. This is far from true in a data-dependent experiment, characterized by under-sampling.

However, many of these limitation were dealt with in a single statistical framework previously developed in out laboratory called weighted spectrum, peptide, and protein model (WSPP) (Navarro *et al.*, 2014a). WSPP model allows a systematic comparison and establishes a validity of null hypothesis at each one of the levels (spectrum, peptide and protein). WSPP model provides a detailed description of the behavior of the technical variance, and by analyzing it independently at the spectrum, peptide, and protein levels, the model is able to capture separately the specific error sources of each SIL and MS method, demonstrating that error distributions are accurately modeled in all cases at the three levels. Last, but not least, it is applicable to any kind of quantitative proteomics experiment and instrument.

5.1.PTM quantification

The quantification of PTMs, although very important to understand the biological processes, is technically challenging. On one hand, PTM-containing peptides are often low abundant. As stated before, the very high dynamic range of protein concentrations in a cell poses serious difficulties for the quantification of PTMs. In addition, it is difficult or impossible to devise a strategy allowing the separation and analysis of all the structurally diverse protein isoforms, because of the number of possible combinations of PTM locations within a protein with several PTMs. Importantly, PTM quantification not only has to cover the proteins change in abundance but also the determination of the modification site.

In general, these enrichment steps enrich a specific type of PTM and/or deplete the big portion of non-modified peptides that impede the PTM analysis. Therefore, this approach creates a need for two separate experiments one for PTM and the other one for unmodified peptides. Different strategies are used to enrich PTMs, for instance, strong cation exchange chromatography is used for phospho or glycopeptides (Lohrig, Sickmann and Lewandrowski, 2011). Some technique also target the physiochemical characteristics of modified peptides, which results in specific retention times in reversed phase (Batth, Francavilla and Olsen, 2014) or Zwitterionic hydrophilic interaction chromatography (Zic-HILIC) (Parker *et al.*, 2013). In some cases, to fractionate the sample more, a combination of affinity chromatography and HILIC is used. On other hand, immobilized metal affinity chromatography (IMAC) (Thingholm and Larsen, 2016a) or Titanium dioxide (TiO2) (Thingholm and Larsen, 2016b) is mainly used for selective enrichment of phosphopeptides (Engholm-Keller *et al.*, 2012). For most PTMs, however, no strategy or protocol for enrichment exists at the moment, and are often either left out of the study, or have to be analysed one by one on protein-by-protein bases (Olsen and Mann, 2013).

Considering the growing interest in the cross talk between different PTMs, there is an urgent need for new methods and tools, which would enable identification and quantification of multiple PTMs simultaneously. Many algorithms (Lu *et al.*, 2007; Cox and Mann, 2008) have provided a promising way for quantification of proteins and can be directly extrapolated to PTMs. For Instance, MaxLFQ a part of MaxQunat However, there is still a need of a generic quantification model for PTMs, which can detect changes in PTM independent of the protein change. A further important consideration is that OS methods have not been previously used to quantify PTM alterations, and a general statistical model for the analysis of data of this kind is currently lacking.

The WSPP model (see also section 5) considers three different variance sources (at the scan, peptide and protein levels), which allows a more accurate modelling of the heterogeneous variance at all levels. This variance heterogeneity at all levels is taken into account by applying weighted averages to quantify scans, peptides and proteins and resolves the under-sampling problem of DDA MS experiments. Along with WSPP model, another model called systems biology triangle (SBT) was developed in our lab, that allows the study of protein coordination by pairwise quantitative proteomics (García-Marqués *et al.*, 2016).

In the past the WSPP model has been applied to modified peptides to analyse the change in the redox state of the cysteine containing peptides (Martínez-Acedo *et al.*, 2012). Analysis was performed in two steps; first, proteins were quantified and peptide variance was estimated with non-cysteine-containing peptides. Furthermore, in a second step, cysteine-peptides were introduced and their deviation was calculated from the corresponding protein. This allows in a single analysis, a reliable detection of quantitative alteration in the peptides independent of the protein change in abundance. However, in second step, when cysteine-peptides are introduced, protein means are recalculated, introducing a bias in protein quantification and reducing the sensitivity to detect modified peptides changes in abundance. Modified peptides derived from proteins detected with less than one non-modified peptide (orphan modified peptides) were excluded from the analysis.

Objectives

Post-translational modifications hugely increase the functional diversity of proteomes. Recent algorithms based on ultra-tolerant database searching (Open search, OS) have opened the way towards unbiased analysis of peptide modifications by shotgun mass spectrometry. However, OS approaches identify only one-half of the modified forms detectable by conventional "closed" searches and do not map the modified residue. Moreover, OS has not been used for the quantification of modified peptides and reliable statistical models for these kinds of experiments are currently lacking.

Mitochondrial heteroplasmy (a condition produced by mitochondrial replacement therapies used in the treatment of infertility and other mitochondrial-linked diseases in humans) has been demonstrated to produce adverse physiological effects. However, the underlying mechanisms are unknown and the biomedical importance of this situation makes it urgent to understand better how these undesired effects are produced.

For these reasons, my PhD thesis is centered on the following general objectives:

1. To develop new algorithms for the comprehensive hypothesis-free identification of all MS-detectable peptide modifications, allowing accurate location of the modified residue.

2. To take further the statistical model for quantitative proteomics previously developed in our laboratory, adapting it for simultaneous quantification of the modified peptidome and the proteome and for systems biology interpretation.

3. To illustrate the performance of the newly developed tools by uncovering the changes in the modified peptidome brought about by mitochondrial heteroplasmy in different mouse tissues.

Material and methods

All the heteroplasmic mice generation and animal handling was performed in the lab José Antonio Enríquez Domínguez. (Functional Genetics of the Oxidative Phosphorilation System (GENOPHOS))

1. Mouse model of heteroplasmy

All animal procedures conformed to EU Directive 86/609/EEC and Recommendation 2007/526/EC regarding the protection of animals used for experimental and other scientific purposes, enforced in Spanish law under Real Decreto 1201/2005. The mice were fed a standard chow diet (5K67 LabDiet).

2. Generation of heteroplasmy mice

Heteroplasmic mice were generated by electro-fusing cytoplasts from conplastic BL/6^{NZB} zygotes to recipient C57BL/6JOlaHsd (BL/6^{C57}) one cell embryos, cultured overnight and transplanted as two-cell embryos into pseudo pregnant Hsd:ICR (CD-1®) females to complete development to term as previously described (Jenuth *et al.*, 1996). To the best of our knowledge, no consensus rule to name heteroplasmic mouse strains exists. Here, we propose the following designation to name heteroplasmic mouse strains: NUCLEAR GENOME-mtCYTOPLASMIC GENOME #1 + CYTOPLASMIC GENOME #2 [i.e., C57BL/6J-mtC57BL/6+NZB, a strain with the nuclear genome of C57BL/6J and the cytoplasmic (mitochondrial) genome of C57BL/6J and NZB]. To simplify we are calling it BL/$6^{C57-NZB}$ along this report. The female heteroplasmic offspring (named BL/$6^{C57-NZB}$) were mated with C57BL/6JOlaHsd males to prevent nuclear genetic drift in our particular mice lines. Only offspring of the established heteroplasmic mice were used.

3. Mice breeding

Heteroplasmic females (BL/$6^{C57-NZB}$) were outcrossed with males BL/6^{C57}. Only females with an initial level of NZB heteroplasmy above 20% were used for colony maintenance.

The mice used in this work were 12-week-old control (C57BL/6JOlaHsd strain) and heteroplasmic males (containing more than one mtDNA in the same cytoplasm, C57BL/6 background). The effect of heteroplasmy on the PTMs of the proteome of different tissues was studied on liver, heart, and skeletal muscle (gastrocnemius) samples. In the last two tissues, the heteroplasmy was stable, while the liver was selected as a control tissue since it spontaneously

selected one of the alternative variants of mtDNA (manuscript in submission For each tissue, biological replicates from different control (N = 3) and heteroplasmic mice (N = 4) were analysed.

4. Benchmarking mass spectrometry dataset

To test the performance of the developed algorithms, we used the publicly available HEK293 dataset (Chick *et al.*, 2015), containing 1.121.149 MS/MS spectra in 24 raw files acquired on a Q-Exactive Orbitrap mass spectrometer. For the bench-marking of site localization through CometPTM we used a synthetic phosphopeptide data set from the Pride database (dataset identifier PXD007058) (Ferries *et al.*, 2017).

5. Preparation of protein extracts

Mice were sacrificed by cervical dislocation. After that liver, heart and skeletal muscle tissues were extracted. 20 mg of each tissue were homogenized in lysis buffer (10mM Tris-HCL pH7.4, 1 mM EDTA, 0.32 M sucrose, 2% SDS) freshly supplemented with protease and phosphatase inhibitors (Roche) and 50 mM DTT, using a MagNA Lyser instrument (Roche). The lysate was boiled for 5 min and cell debris were removed by centrifugation.

6. Protein digestion, peptide labelling and fractionation

Proteins were treated with 50 mM iodoacetamide (IAM) and digested with trypsin using the Filter Aided Sample Preparation (FASP) digestion kit (Expedeon) (Wiśniewski, Zielinska and Mann, 2011) according to manufacturer's instructions. Dried peptides were labeled using 10 plex-TMT reagents according to manufacturer's instructions (Thermo Fisher Scientific), desalted on OASIS HLB extraction cartidges (Waters Corp.) (Leyfer and Weng, 2005), separated into 7 fractions using the high pH reversed-phase peptide fractionation kit (Thermo Fisher Scientific) and dried-down before MS analysis.

7. LC-MS analysis

Each fraction of the labelled peptide samples were analysed using an Easy Nano-flow HPLC system (Thermo Fisher Scientific) coupled via a Nano-electrospray ion source (Thermo Fisher Scientific, Bremen, Germany) to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). C18-based reverse phase separation was used with a 2-cm trap column and a 50-cm analytical column (EASY column, Thermo). Peptides were loaded in buffer A (0.1% formic acid (v/v)) and eluted with a 240 min linear gradient of buffer B (80%

acetonitrile, 0.1% formic acid (v/v)) at 200 nL/min. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS using a top 15 method. MS spectra were acquired in the Orbitrap analyser with a mass range of 400–1500 m/z and 60,000 resolutions. HCD fragmentation was performed at 27 of normalized collision energy and MS/MS spectra were analysed at 60,000 resolutions in the Orbitrap.

8. Database search

Unless indicated otherwise, all searches were performed using Comet release 2016 (Eng, Jahan and Hoopmann, 2013b; Eng *et al.*, 2015). Using trypsin digestion with 1 missed cleavages (unless otherwise specified) and fixed Cys carbamidomethylation (57.021464 Da). For heteroplasmic mice data, TMT labelling at N-terminal end and Lys was also considered as a fixed modification (229.162932 Da). Fragment ion tolerance was 0.02 Da, 0 mass offset. Precursor tolerance type and isotope error were set to 1. Precursor charge range was 2-4, maximum precursor charge 5 and maximum fragment charge 3. Only y- and b-ions were used for scoring.

Closed searches (CS) were performed at 5-ppm precursor ion tolerance, using three dynamic modifications: Met oxidation (15.994915), Asn and Gln deamidation (0.984016) and Ser and Thr phosphorylation (79.966331). Peptide identification from MS/MS data was performed using the probability ratio method (Martínez-Bartolomé *et al.*, 2008). False discovery rates (FDR) of peptide identifications were calculated using the refined method (Navarro and Vázquez, 2009; Bonzon-Kulichenko *et al.*, 2015); 1% FDR was used as the default criterion for peptide identification. Open searches (OS) with Comet and Comet-PTM were performed in the same conditions as CS, except that precursor ion tolerance was set to 500 Da.

Results

In the theme of this thesis, to overcome and to highlight the issues regarding PTM identification and studies build around PTMs, the results are divided into three sections: development, benchmarking and application. In development section, a detailed explanation of novel algorithms and tools developed is provided. In benchmarking, the performances of developed tool is analysed and discussed thoroughly. To show the impact of these newly developed algorithms, the suit of these bioinformatics tool was applied on a biological model of mitochondrial heteroplasmy and biological conclusions are discussed concerning PTMs.

1. Development of novel bioinformatics tools

1.1. Comet-PTM

CometPTM was developed (in collaboration with Spiros Michalakopoulos and Marco Trevisan-Herraz) by modifying the open-source database search engine (Eng et al., 2015). For every sequence candidate Comet-PTM calculates the difference between theoretical and experimental precursor mass (ΔMass), and adds up this mass iteratively to each one of the amino acid masses in the peptide sequence, calculating a Xcorr score in each one of the possible modified forms of the peptide (Figure 4). The selected candidate is the modified peptide form that produces the highest Xcorr. This design allows Comet-PTM to reach the score that would have been obtained by performing a targeted CS with the same modification in the same position (Also with OS). Note that the scores are not exactly identical, since CS uses the theoretical mass of the modification and Comet-PTM estimates it from the difference between the precursor mass and the theoretical mass of the non-modified peptide, and experimental errors on this estimate may affect fragment matching. This effect is, however, small when low ppm precursor mass accuracies are used (Figure 5 A and B). Comet-PTM has a user-selectable option of scoring also the non-modified peptide sequence (even when ΔMass is different from zero), to take into account labile modifications (Kong et al., 2017).

Comet-PTM was developed to take full advantage of the multi-thread design of Comet. Comet-PTM used less than 4 hours to perform a 500 Da-wide open search of 16 LC-MS runs, containing an average of 44,390 MS/MS spectra each, using a computer cluster with 16 nodes, where each node is built of 2 x Intel Xeon E5-2695v2 at 2.40 GHz and contained 46 threads/124 gigabyte.

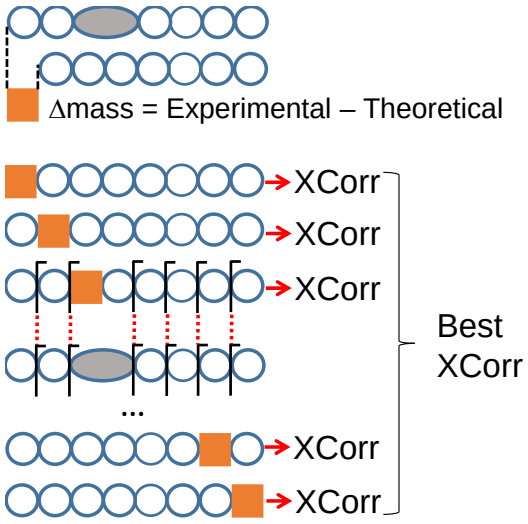


Figure 4: Schematic representation of comet-PTM algorithm. Comet-PTM firstly calculates the difference between the mass of the candidate and the mass of the precursor ion detected by the mass spectrometer (ΔMass). ΔMass is then iteratively added to each amino acid in the peptide sequence and the position that yields the best score is selected as the correct match.

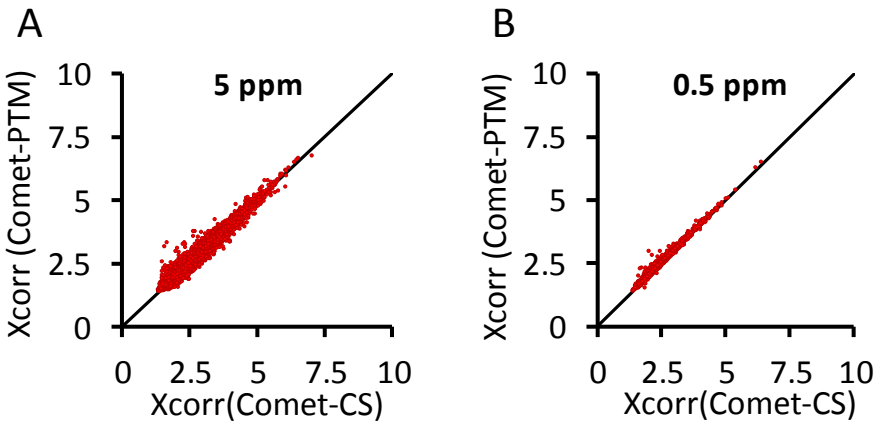


Figure 5: Peptide identification performance of Comet-PTM (A-B). Comparison of scores obtained from Comet-PTM and CS in the population of PSM that produced the same match with the two engines. A match was considered identical when the peptide sequence was the same and the difference between ΔMass obtained by Comet-PTM and the theoretical mass of the modification selected in CS was within 5 ppm (A) or 0.5 ppm (B). Note that the scores were practically identical, and the dispersion around the identity line was diminished when the tolerance decreased; this demonstrates that the small differences in the score are a consequence of the error in the estimation of ΔMass and not in the design of the score in Comet-PTM.

1.2. Systematic Hypothesis-free Identification of modifications with controlled FDR based on ultra-Tolerant database Search (SHIFTS)

SHIFTS is a program that was develop to identify peaks in frequency distribution of modified peptides on the basis of the mass change/shift of their modification (ΔMass distribution), to assigns PSM to peaks and to calculate FDR for peptide identification (Figure 6). SHIFTS uses as input the files obtained from Comet-PTM search.

1.2.1. Mass recalibration

SHIFTS firstly recalibrate precursor peptide masses independently in each raw file. Recalibration was performed on a population of non-modified peptides with a very high score (user selectable; recommended values are those yielding 0.1% global FDR or lower). The selected population were assumed to be true identifications and are used to calculate the systematic mass error (median deviation in m/z scale), which is assumed to be constant in each raw file. From these data, SHIFTS also calculates the standard deviation of the mass error (σ_M) using the median absolute deviation (MAD) method (Figure 6).

1.2.2. Peak identification

Recalibrated ΔMass values were binned using 0.001 Da bins to construct the ΔMass distribution. The distribution was smoothed using the median of a 7-point sliding window and then peak apexes were detected as downward zero-crossings in the first derivative of the smoothed curve. Peak widths were similarly calculated as the zero-crossing points of the second derivative; in the current version of SHIFTS they are computed only for informative purposes.

1.2.3. Peak assignation

By default SHIFTS assigns a PSM to the closest ΔMass peak if the mass deviation of the PSM from the peak falls within$3\sigma_M$, so that approximately 99% of PSM in each peak are assigned. This value can be user adjusted. PSM not assigned to peaks were considered as orphan PSM. The use of MAD to assign a PSM to peak was driven by simple hypothesis; that the distribution of ΔMass at any charge or any range of ΔMass follows the same distribution of "True" non-modified peptide distribution. Hence, the deviation of very high scoring non-modified PSM was measured as a MAD (Figure 7).

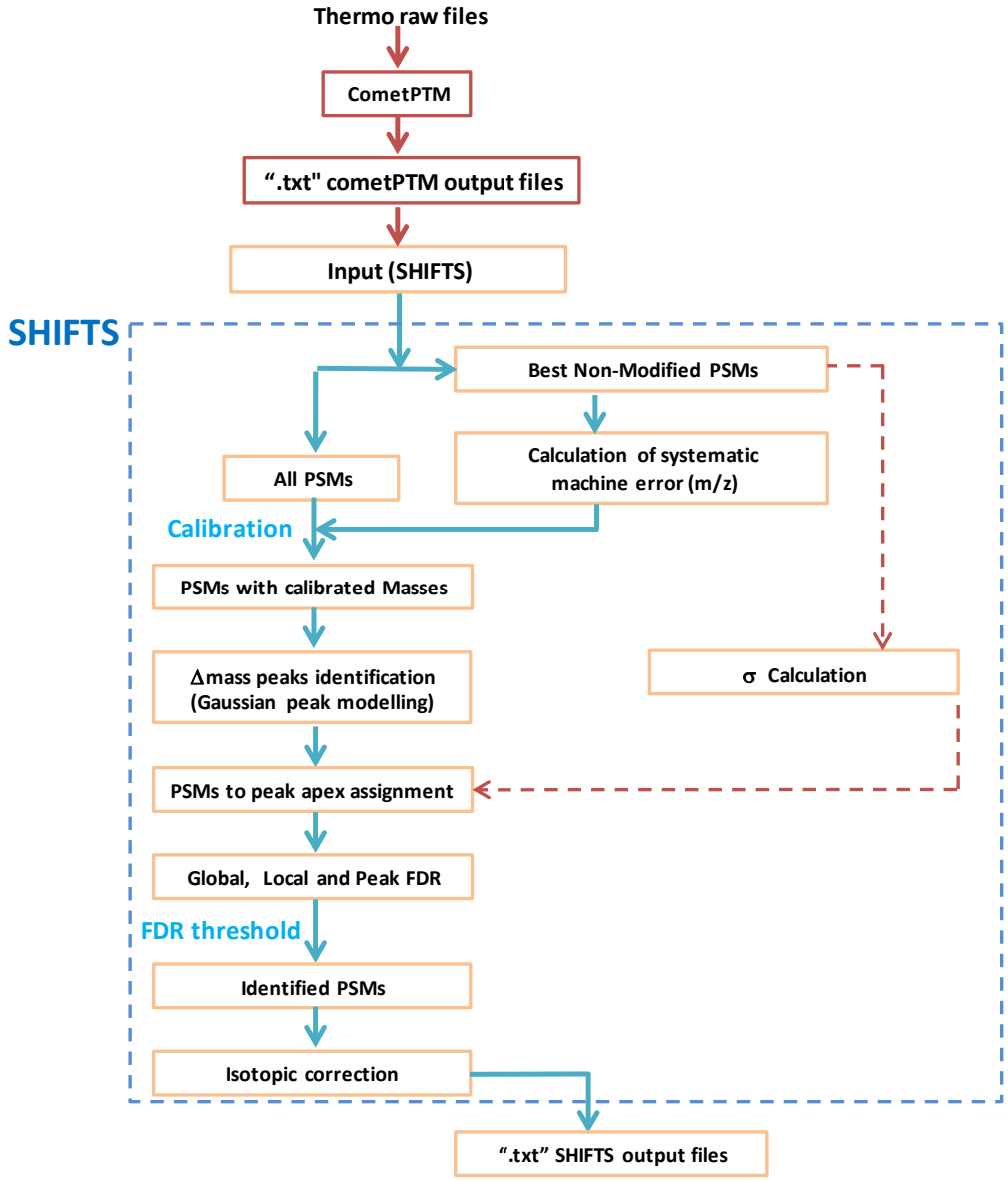


Figure 6: Schematic representation of SHIFTS. The workflow depicts how the output from Comet-PTM is processed, including mass recalibration, peak detection and FDR calculation

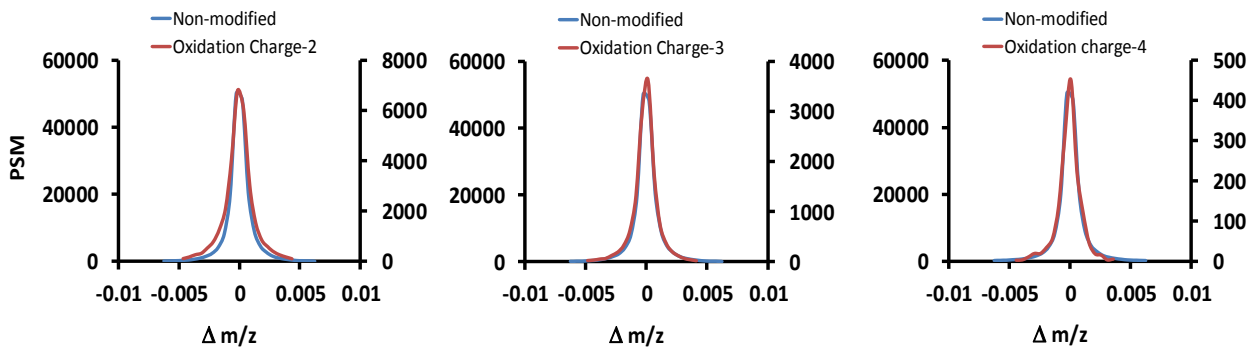


Figure 7: distribution of Oxidation in different charge state. Plots represents distribution of "True" non-modified peptides, which was used as MAD to assign PSMs to peak apex.

1.2.4. FDR calculation

SHIFTS calculates the FDR of identification using a conventional target/decoy strategy using the corrected Xcorr score (cXcorr) (Keller *et al.*, 2002; Choi, Fermin and Nesvizhskii, 2008). A *global FDR* was calculated for each PSM as the ratio of the number of decoy PSMs to the number of target PSMs having a cXcorr equal or higher. Decoy peptides matched by Comet-PTM were observed to be almost as abundant as target peptides in the negative ΔMass region below the peak corresponding to neutral loss of Gly (Figure 8A), where ΔMass peaks were mostly produced by neutral loss of amino acids. For this reason, the global FDR was only calculated in the ΔMass region above -56 Da (Figure 8A). All the PSM are required to have FDR lower than the global FDR, without exception.

In addition, local FDR filters are also applied. Some ΔMass peaks were observed to contain an unusually high number of decoy PSMs. To avoid matching false positive target PSM in these peaks, SHIFTS also calculates a *peak FDR* counting up the number of decoys and target PSM assigned to each peak and these PSM are required to pass the peak FDR filter in addition to the global FDR filter (Figure 8C). Note that peak FDRs are often very low suggesting that the majority of PSM in these peaks are true, even when they have a low cXcorr. This happens because the probability of finding a decoy PSM in a peak by chance alone is extremely low. SHIFTS avoids matching these low scoring target PSMs by applying the global FDR filter.

To apply a local filter to PSM which are not assigned to ΔMass peaks, e.g. to orphan PSM, SHIFTS models the periodic mass distribution of decoy PSM into ~1 Da-bins centered at the regions where ΔMass values concentrate, and calculates a *local FDR* by counting up decoy and target PSM in each one of these regions (Figure 8B). The local FDR filter is applied 0to orphan PSM in addition to the global FDR filter. Default values for peptide identification were 1% for peak and local FDR; 5% for global FDR.

1.2.5. Isotopic correction

SHIFTS also perform a simple isotopic correction to minimize miss assignations of the correct monoisotopic peak of the precursor. When two PSM having the same sequence are encountered having a ΔMass difference within 1 ppm of the mass difference expected for either one or two ^{13}C or one ^{34}S, that of the lightest one substitutes the ΔMass of the heaviest precursor.

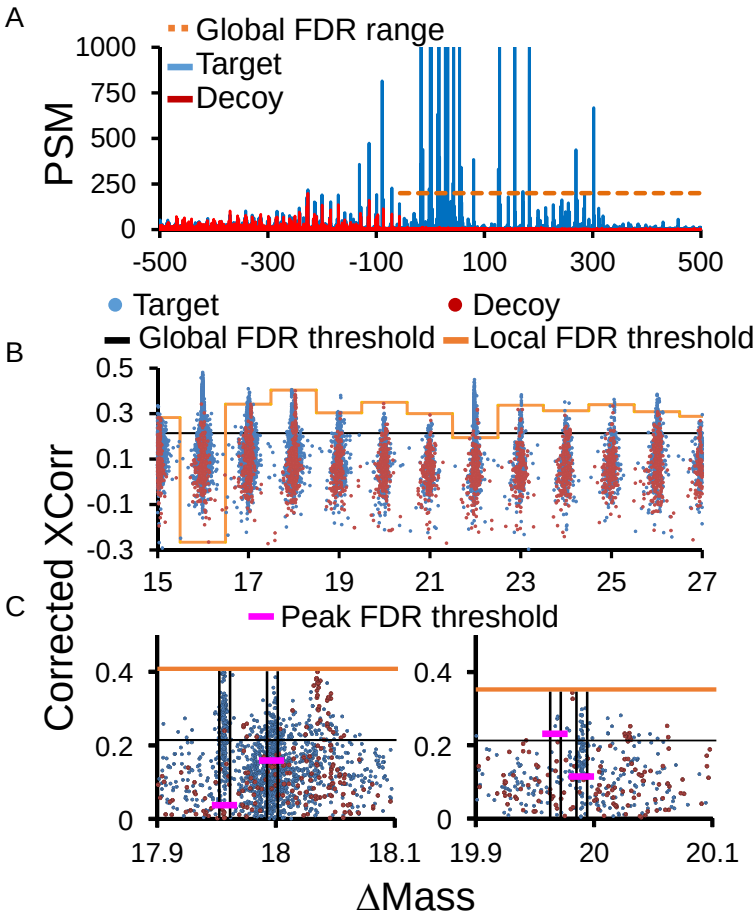


Figure 8: Distribution of target/decoy PSMs and FDR calculations. (A) ΔMass distribution of target and decoy PSMs and range of application of the global score threshold. The vertical scale is enlarged to show the distribution of decoy PSMs. (B) Local score thresholds applied in each ~1 Da bin in the ΔMass distribution(C) Peak score thresholds applied to the peaks detected by SHIFTS in the ΔMass distribution.

1.3.PtmSticker: connecting dots and translating numbers to meaning

PtmSticker is an algorithm that enables semi-supervised annotation of peptide modifications. In the Comet-PTM output, PtmSticker searches the ΔMass values against the Unimod database with a user-defined tolerance (in PPM range), taking into account the modified amino acid. In addition, the algorithm also takes into account the preceding and the consecutive residues of the initially modified residue and compares them with the list of amino acids that could be subjected to the modification according to Unimod. Upon failing to match the Δmass and the residue in UniMod, it checks the residues, containing modifications that were fixed in the database search. Residues with fixed modifications were considered with and without the fixed modification; when not indicated, the amino acid contains the fixed modification (e.g., C_oxidation means oxidation of carbamidomethylated Cys and K_oxidation, oxidation of TMT-labelled Lys). ΔMass

values that could not be matched were tentatively tested assuming one ^{13}C miss assignment of the monoisotopic mass of the precursor and as combinations of two modifications from a list of the most abundant modifications found in the corresponding proteome (Figure 9). Unexplained ΔMass values were termed as unknown.

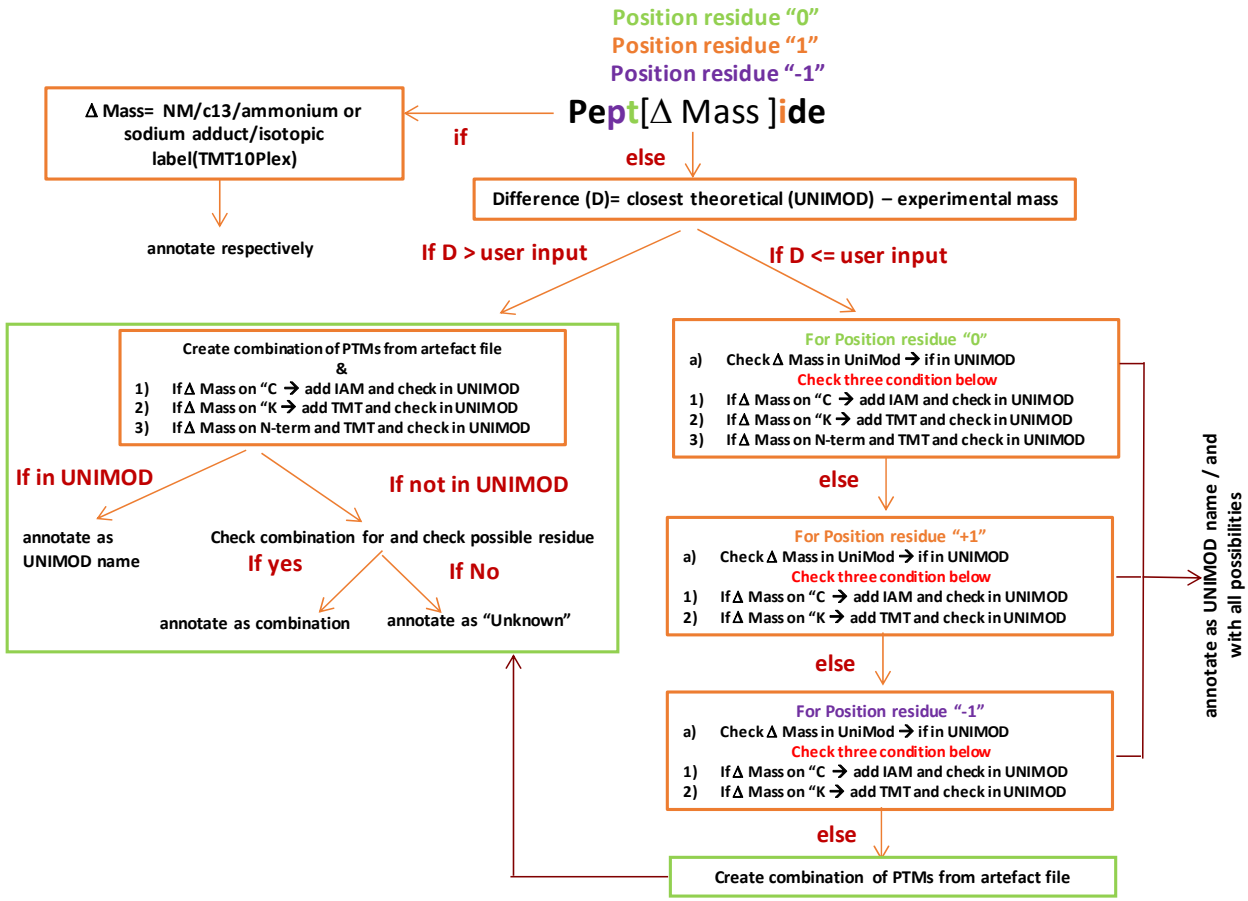


Figure 9: a workflow showing the steps in PtmSticker. PtmSticker, firstly assigns and labels the ΔMass of non-modified, c13, ammonium, sodium adduct and isotopic label (for instance, TMT10Plex) respectively. Further, it checks if the difference is smaller than user input, if so, it takes in account modified residue and check for it in UniMod database. Upon not finding it in UniMod at 0th position, it considers preceding and consecutive amino acids. PtmSticker takes in account the static modification and labelling mass if specified. As a last resort in cases where it fails to find in UniMod, PtmSticker creates a set of combination (from user provided list), check them in UniMod, and labels them accordingly.

1.4.PTM quantification as complement to WSPP model

The quantitative information from TMT reporter intensities was integrated from the spectrum level to the peptide level and then to the protein level on the basis of the WSPP model (Martínez-Acedo *et al.*, 2012; Navarro *et al.*, 2014b) using the Generic Integration Algorithm (GIA) (García-Marqués *et al.*, 2016).

The new algorithm "SanXotGhost" was developed and later also was include into GIA workflow for the statistical model the quantitative values of modified peptides as part of the automated workflow (Figure 10A). Before PTMs can be quantified, another tool was developed called "Trilogy". Trilogy generates peptide-to-protein relations tables in which peptides are tagged as non-modified, modified or orphans (e.g., when the protein is detected only with modified peptides). This allows user-selectable calculation of protein averages and peptide variances using only the unmodified peptide forms. Hence, it is possible to detect modified peptides that significantly deviate from the weighted average of the unmodified peptides from the same protein

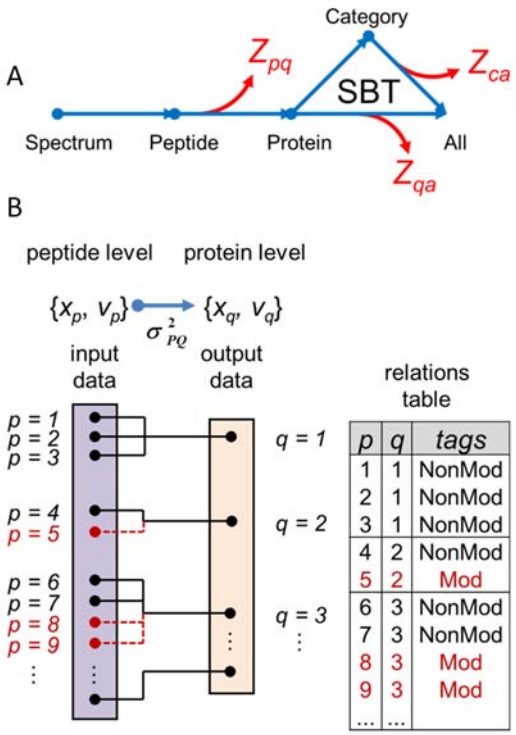


Figure 10: Quantification workflow (A, B). (A)Scheme of the integrative workflow used to quantify modified peptides, proteins or categories. Each arrow represents a step performed with the generic integration algorithm. Standardized $\log_2$-ratio values at the peptide level (Zpq) are obtained from the peptide-to-protein integration. The algorithm provides corrected peptide values by taking account of the corresponding protein changes. The same workflow is used for protein quantification and systems biology analysis of coordinated protein responses with the SBT model. (B) Scheme of the generic integration algorithm (GIA) adapted for the analysis of modified peptides in the peptide-to-protein integration step. The relations table include tags that can be used to indicate that modified peptides are excluded from the computation of protein averages.

Briefly, for each sample i the values $x_{qps} = \log_2 S_i/C$ were calculated, where S_i is the intensity of the TMT reporter corresponding to sample i in the MS/MS spectrum s coming from peptide p and protein q, and C is the average intensity of all the TMT reporters from the control samples, which is used as a common reference. The $\log_2$-ratio of each peptide

(x_{qp}) was calculated as the weighted average of its spectra, the protein values (x_q) were the weighted average of its peptides, and the grand mean ($\bar{x}$) was calculated as the weighted average of all the protein values (Navarro *et al.*, 2014b). The statistical weights of spectra, peptides and proteins (w_{qps}, w_{qp} and w_q, respectively) and the variances at each one of the three levels ($\sigma_{\mathbf{S}}^2$, $\sigma_{\mathbf{P}}^2$, and $\sigma_{\mathbf{Q}}^2$, respectively), were calculated as described(Navarro *et al.*, 2014b). For each modified peptide, the standardized variable (z_{pq}) was calculated as

$$\boldsymbol{z_{pq} = \left(x_{qp} - x_q\right)\sqrt{w_{pq}}\sqrt{\frac{n_p}{n_p - 1}}, \qquad n_p > 1}$$

Where n_p is the number of non-modified peptides with which the corresponding protein was quantified. z_{pq} Expresses the deviation between the peptide $\log_2$-ratio and the corresponding protein value in units of standard deviation. Similarly, orphan peptides can be quantified using grand mean of proteins.

$$\boldsymbol{z_{pq} = \left(x_{qp} - x\right)\sqrt{w_{pq}}\sqrt{\frac{n}{n - 1}}}$$

Where *n* denotes the number of proteins identified in experiment, and *x* is protein grand mean.

2. Benchmarking of algorithms and tools

2.1. Comet-PTM enables comprehensive identification of peptide modification

To test the performance of Comet-PTM, the results were compared with those obtained with OS and with CS using 3 common variable modifications. As expected, OS assigned higher scores to a large population of peptide-spectrum matches (PSM) containing modifications not included in the CS list of variable modifications (Figure 11A, left). However, CS assigned a higher score than OS to a large population of PSMs with modifications included in the CS list, because the modifications affected the matching of fragment ions, decreasing the OS-assigned score (Figure 11A left). This effect reduced the identification performance of OS to around 50% of the PTMs identified by CS (Figure 11B), confirming previous results(Chick *et al.*, 2015). In contrast, scores obtained with Comet-PTM matched or exceeded OS (Figure 11A, centre).

In addition, Comet-PTM identified the same ΔMass peaks as OS (Figure 11C); however, most of these peaks contained about 2-fold more PSMs (Figure 11C-H), reflecting the superior performance. As expected, this effect was not observed when the modification did not affect the fragmentation series. Thus, the ^{13}C peak, produced by errors in assignment of the monoisotopic precursor mass, was observed with exactly the same number of PSMs (Figure 11I). The number of PSMs was also similar for modifications in N-terminal position of the peptide, which only affect b-series and has a negligible effect on identification by HCD fragmentation (Figure 11J, K).

Comet-PTM produced scores similar to or higher than those obtained with CS (Figure 11A, right), except for a small population of peptides for which CS found 2 or more modifications (Figure 11A, orange dots). This effect was not due to differences between CS and Comet-PTM scores (Figure 5A and B). Therefore, and unlike OS, Comet-PTM had a similar identification performance to that of CS for the preselected modifications (Figure 11B). In several instances, Comet-PTM correctly located an oxidation on Trp, Pro or Tyr that CS wrongly assigned to oxidation on Met, the predefined variable modification residue in CS (Figure 12 A, B, C).

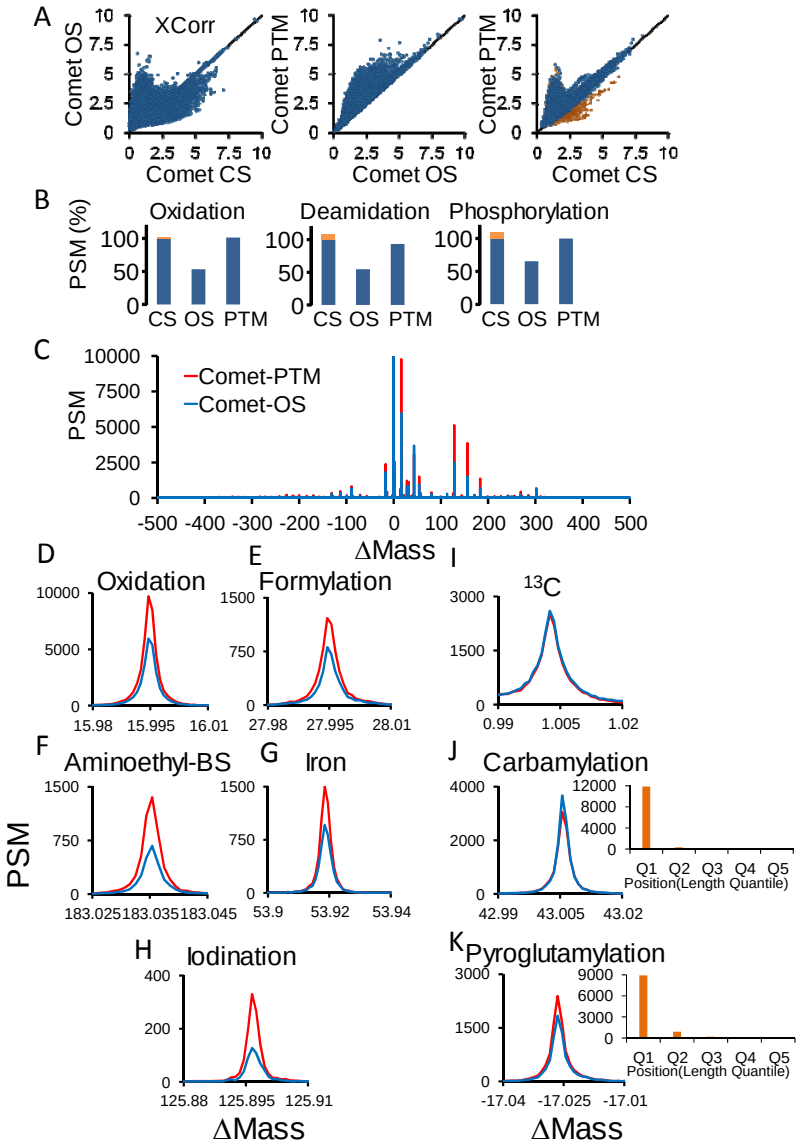


Figure 11: **overview of peptide identification and CometPTM performance.** (A)The HeLa dataset from the original OS article was searched using Comet in closed search (CS) mode with 3 variable modifications (Met oxidation, Asn and Gln deamidation and Ser and Thr phosphorylation), with Comet in OS mode, or with Comet-PTM (500 Da tolerance in the later 2 cases). The scores obtained for the same spectra in the different searching conditions are compared. Yellow points in the rightmost graph are PSMs that match peptides with more than one modification in CS mode. (B)Identification performance of OS and Comet-PTM relative to CS in the populations of peptides modified by oxidation, deamidation or phosphorylation. The number of PSMs was obtained after filtering by the score threshold corresponding to 1% FDR in the CS. PSMs matching peptides with more than one modification in CS are indicated by orange bars. (C) Frequency distribution of PSMs obtained by OS and Comet-PTM as a function of ΔMass. (D-K) Details of the frequency distribution around the indicated modifications. The charts to the right of (J) and (K) show the distribution of PSMs as a function of the quantile position of the modification in the peptide sequence assigned by Comet-PTM

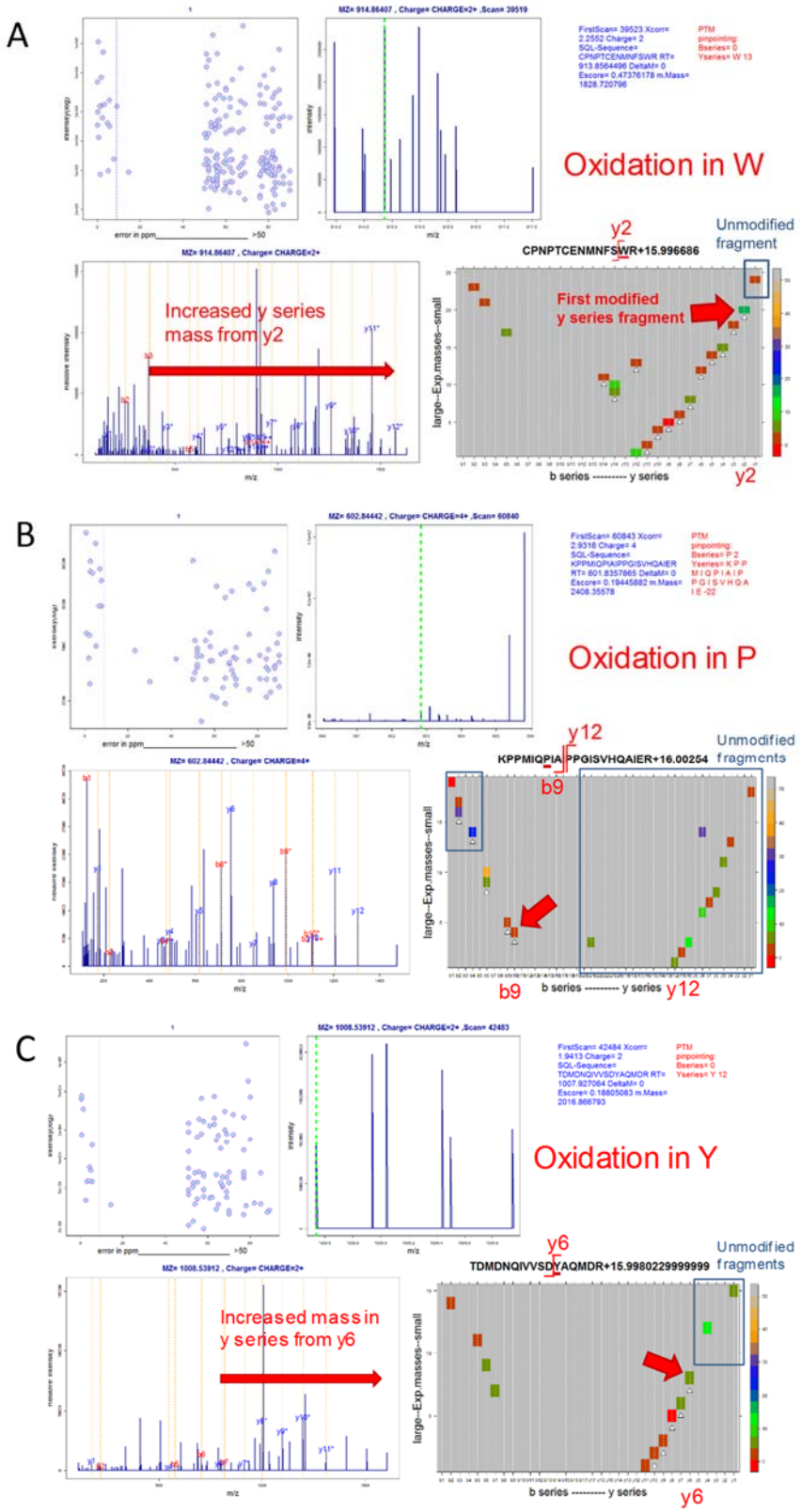


Figure 12: Vseq analyses of 3 peptides containing oxidations incorrectly assigned to Met by CS. Comet-PTM correctly assigns the oxidation to Trp (A) Pro (B) or Tyr (C).

2.2. Comet-PTM detects the location of modifications in the peptide sequence

Peptides were identified after Comet-PTM searches with SHIFTS, an algorithm that detects the peaks in the ΔMass distribution and controls the peptide false discovery rate (FDR) through a conservative, 3-layered approach (Figure 8A-C; see also result section 1.2). From the output of a search against a concatenated target-decoy database, SHIFTS calculates a global score threshold to control FDR in the population of PSMs with ΔMass > -56 Da (Figure 8A). A local score threshold is also defined to control FDR separately within each of the ~1 Da-bins in the ΔMass distribution (Figure 8B). Finally, a peak score threshold is calculated to control FDR separately within each peak detected in the ΔMass distribution (Figure 8C). A PSM is considered to be correctly identified when its score is above the global and peak thresholds; when a PSM does not form part of a peak, instead of the peak threshold the local threshold is used. This conservative approach allows full control of FDR, avoiding any bias due to the specific behaviour of certain kinds of peptide modifications that may be more prone to match decoy sequences.

An advantage of Comet-PTM over existing OS approaches is that it automatically assigns the modification to the residue in the peptide sequence that produces the best score and therefore is the modified position that best explains the fragmentation data. Assigning modifications to specific residues is considered a much less reliable process than identifying peptides, partly because there is frequently insufficient information to determine the exact modified residue (Chalkley *et al.*, 2008). To estimate the accuracy of Comet-PTM assignation of a modification to the correct site, we filtered correct identifications using the conservative method described above with a 1% FDR threshold and then calculated the fraction of PSMs in which the mass shift was located in amino acids predicted to harbour well-known modifications. Oxidation was located to Met, Trp or Pro in 82% of cases, and deamidation to Asn, Gln or carbamidomethylated Cys in 86% of cases (Figure 13A). Considering 6 well-known modifications, the overall accuracy of Comet-PTM was estimated at ~85%. To test more specifically the performance of Comet-PTM for the correct assignation of phosphorylation sites, we analysed a dataset obtained from synthetic phosphorylated peptides (Ferries et al., 2017). Comet-PTM assigned the correct position with 81% accuracy, and in 94% of the cases, the modification was in the correct or in an adjacent position (Appendix Table 1).

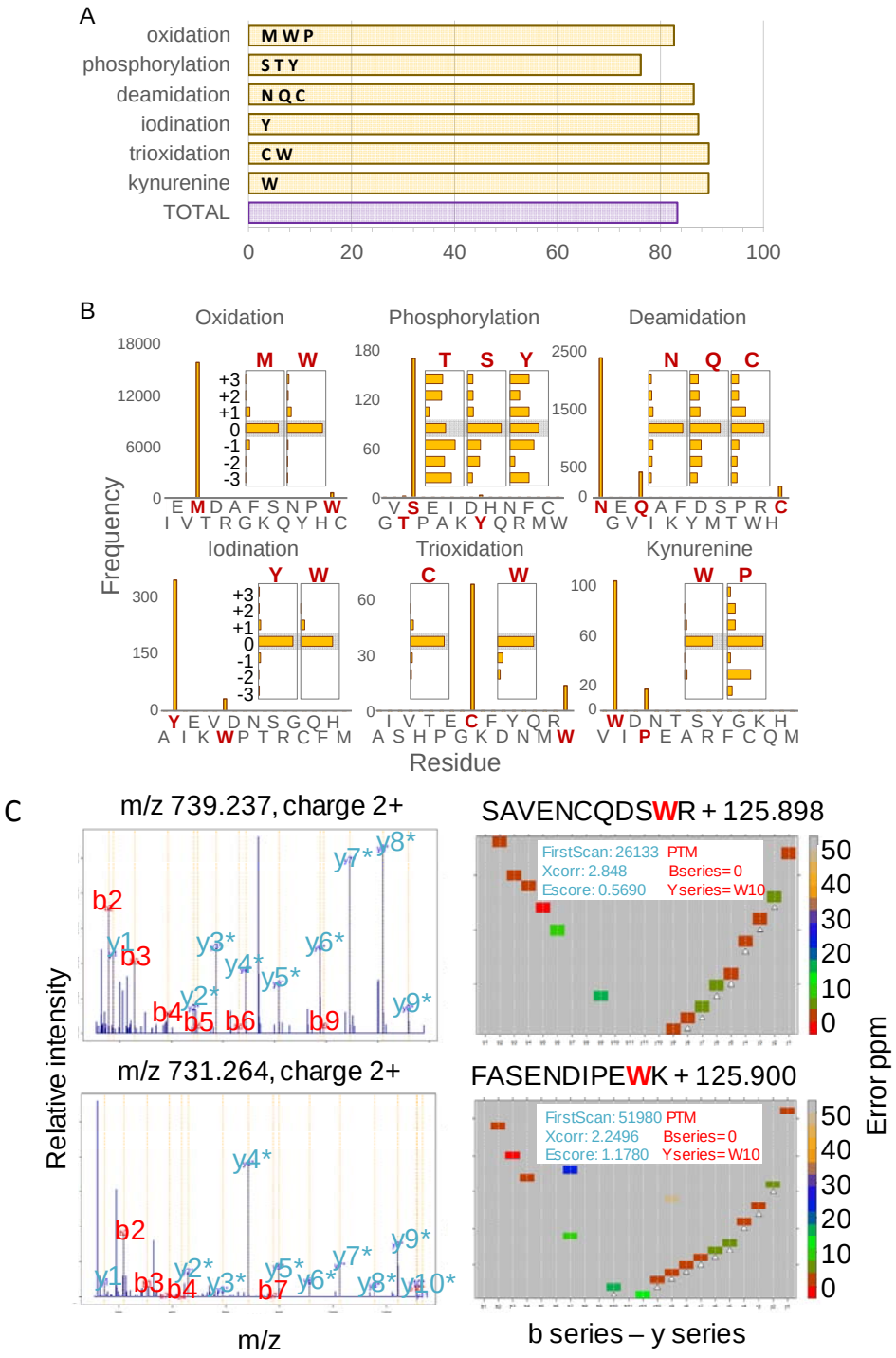


Figure 13: Identification of modified peptides and location of the modified site. (A) Percentage of successful assignment of the indicated modification to the indicated residues. (B) Inset horizontal bar graphs plot the frequency distributions of PSMs assigned by Comet-PTM to the indicated amino acids and modifications. Vertical bar graphs: frequency of PSMs assigned to the indicated amino acids after substracting the average frequency of the 3 previous and the 3 subsequent positions. Residues that accumulate counts in the histogram at position 0 are labelled in red. (C) Vseq results for 2 representative peptides containing iodinated Trp. Both the MS2 spectra (left panels) and the V-shaped heatmap distributions for the main fragmentation series (right panels), demonstrate the location of the modification in Trp.

We then analyzed which amino acids were assigned to known modifications with a higher frequency than expected by chance. Among the most frequent, we only found well-known modifications like oxidized Met and Trp, deamidated Asn, Gln and carbamidomethylated Cys, phosphorylated Ser, trioxidized Cys and Trp and kynurenine-modified Trp (Figure 13B). Some Pro modifications were also found with the ΔMass of kynurenine-modified Trp (Figure 13B), but all of them corresponded to homologous peptides that contained the Pro > Thr substitution, which has the same ΔMass (Figure 14), and that were assigned the same score.

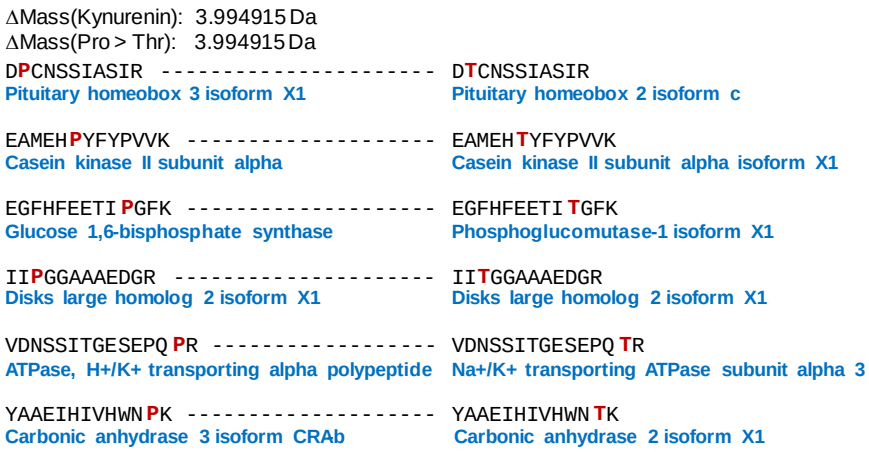


Figure 14: peptides containing a Pro > Thr substitution assigned as a kynurenin modification in Pro. Peptide sequences together with their corresponding proteins are listed, highlighting (in red) the position of Pro and Thr residues.

Finally, Trp iodination was also clearly detected above background (Figure 13 B). This modification was unexpected because although Trp halogenases have been described in bacteria (van Pée and Patallo, 2006) they are not present in mammals, and Trp iodination has not been described before as a chemical artifact. The majority of peptides containing iodinated Trp contained neither Tyr nor His, the 2 amino acids known to react with iodine. A careful inspection of their fragment spectra confirmed the presence of y-fragments that could only be explained by Trp being the modified residue (Figure 13 C). In addition, the only modification described in Unimod that matched the observed mass shift corresponded to iodination. From all these data, we concluded that Trp was iodinated in these peptides, most probably as a side-chain reaction of the iodine produced from the Cys-alkylating reagent iodoacetamide. This finding further highlights the accuracy with which CometPTM locates the site of modification and suggests that this OS engine may be useful for detecting novel peptide modifications in specific amino acids.

2.3. A single integrated statistical framework allows quantification of the proteome and of the modified peptidome

Quantification algorithm includes a peptide-to-protein integration step that quantifies protein values from the unmodified peptide forms and then computes the standardized log2-ratio of the modified peptides with respect to these protein values (Z_{pq}) (Figure 10B). This makes it possible to detect modified peptides whose behaviours deviate significantly from those of the unmodified peptides from the same protein. When applied to a biological model, the Z_{pq} distributions of the unmodified and modified peptide forms precisely followed the expected null hypothesis distribution in different tissues from two mouse strains (Figure 15).

In absence of changes, the distributions of z_{pq} followed very closely the normal distribution $N(0,1)$ (Figure 10C), validating the accuracy of the model. These results convincingly demonstrate that the quantitative behaviour of the entire population of modified peptides can be modelled accurately, providing a robust statistical framework within which to analyse abundance changes in high-throughput experiments. This allowed detection of significant changes in certain modified peptides against the null hypothesis in specific situations (Figure 15, lower left).

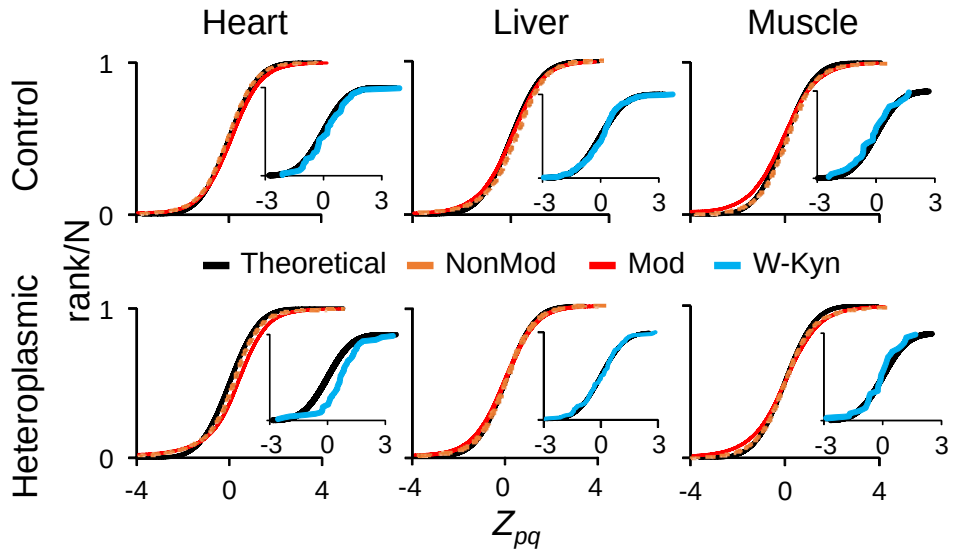


Figure 15: Distributions of *Zpq* values for nonmodified (Non-Mod) and modified peptides (Mod). These data come from the analysis of mouse tissues in a model of mitochondrial heteroplasmy; each graph correspond to a tissue from an individual animal. The inset plots show the *Zpq* distributions for peptides containing the kynurenine Trp modification. Note that the distribution is displaced to the right in the hearts of heteroplasmic mice, indicating increased abundance.

The same workflow used for peptide quantitation allowed the analysis of protein abundance changes and the characterization of functional categories that were affected by the coordinated action of proteins, using the Systems Biology Triangle (SBT) model (García-Marqués *et al.*, 2016) (Figure 10A). This statistical design thus allowed a full description of the alterations in the modified peptidome in the global context of changes in the proteome. Significant abundance changes of modified peptides were detected by Student's t-test comparing the z_{pq} values of samples from heteroplasmic mice (N=4) with those of control mice (N=3) (Appendix Table 2)

3. Application of developed tools

3.1. Heteroplasmy produces protein alterations consistent with a mitochondrial dysfunction in the heart

We used Comet-PTM and associated tools to study the molecular impact of mitochondrial heteroplasmy on the proteome and the modified peptidome in heart, liver and skeletal muscle from 12 week-old mice. This age was considered the most appropriate to study molecular alterations before more extensive damage was produced as a consequence of heteroplasmy. SBT model analysis of the quantitative protein data from the heart revealed a coordinated decrease of proteins related to muscle function and the oxidative phosphorylation system (OXPHOS) and a coordinated increase of tubulins, myosins and proteins involved in vesicle trafficking. The muscle-function proteins included KCRS (mitochondrial creatine kinase) and VDAC3, which are key suppliers of phosphocreatine to the sarcomere, and SGCG and KGP1, which are essential for sarcomere contraction (Appendix Table 2). These alterations in heteroplasmic mice are highly consistent with the decreased ATP synthesis and the abnormal increase in phosphocreatine/ATP ratio in the heart and with the increase in plasma creatine kinase we have reported in this animal model (unpublished data), and provide evidence that heteroplasmic cardiac mitochondria have a compromised ability to supply energy to the sarcomere. Among the increased vesicle trafficking proteins, STXB3 and VAMP2 are implicated in the insulin-dependent movement of GLUT4 from inner vesicles to the plasma membrane. This finding also agrees with the higher glucose uptake we detected in heteroplasmic hearts, pointing to a shift toward glycolytic metabolism in order to compensate the reduced phosphocreatine supply from the OXPHOS. The coordinated decrease in blood proteins and the increase in heart cytoskeletal proteins in heteroplasmic animals suggest a homeostatic effort to maintain cardiac cell structure.

Heteroplasmy did not affect OXPHOS protein levels in liver, but did produce a coordinated increase of DHA signaling pathway proteins (Figure 16), including AKT2, BID and B2CL1 (Appendix Table 2). AKT2 plays an essential role in the progression of the inflammatory response and apoptosis (López-Carballo *et al.*, 2002); consistent with this finding, we detected signs of inflammation in the liver of heteroplasmic mice with aging (unpublished data). Heteroplasmic liver also showed coordinated increases in a group of myosins and phosphatidylinositol pathway proteins, probably reflecting a homeostatic response. These

data thus indicate that at 12 weeks of age, heteroplasmy produces a decrease in OXPHOS proteins in the heart, probably reflecting mitochondrial dysfunction, but the same effect is not seen in liver. In heteroplasmic muscle, we detected a coordinated decrease of proteins regulating cell shape, such as tubulins and microtubule proteins (Figure 16). On the other hand, the thick-filament myosins, as well as calcium transporting proteins, were increased, suggesting the induction of compensatory mechanisms. However, we found no evidence of altered mitochondrial protein levels in muscle at this age.

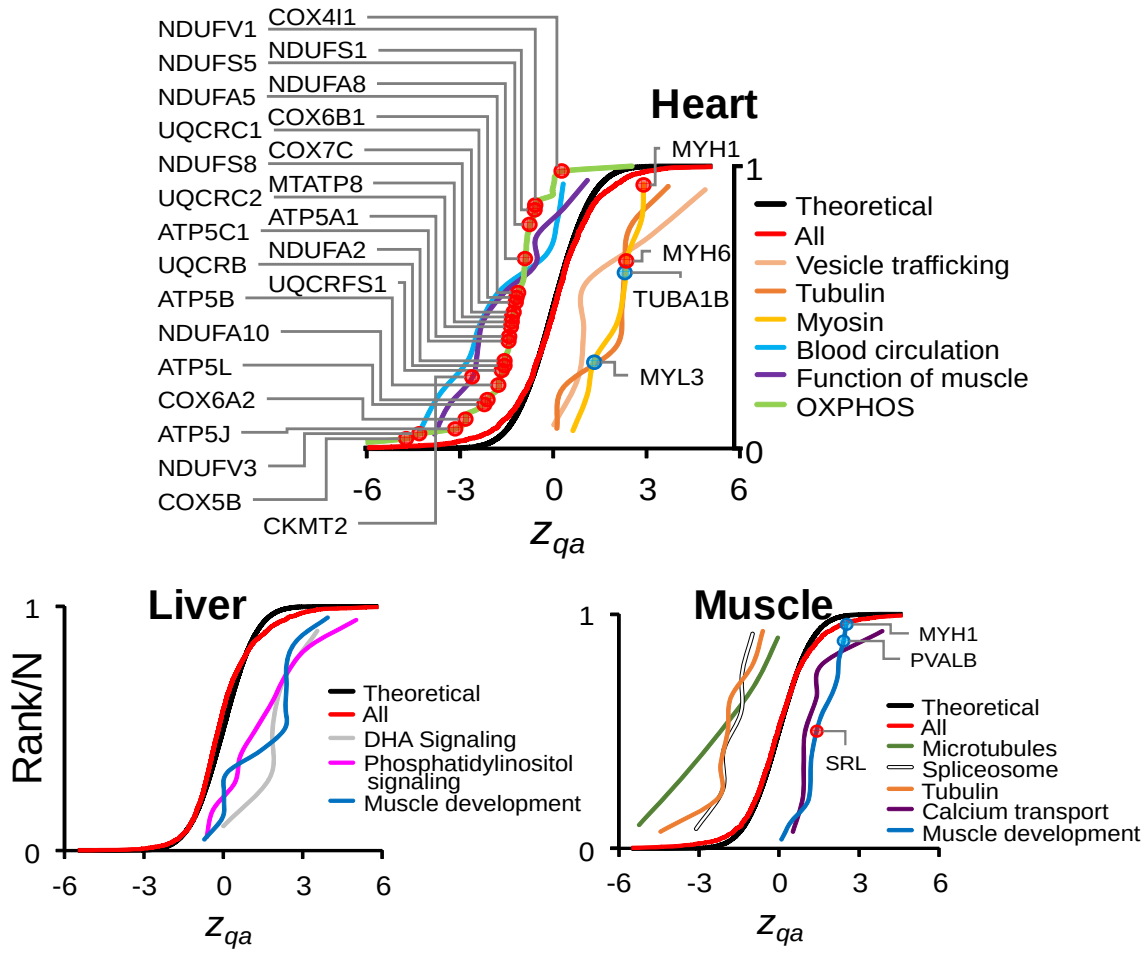


Figure 16: Characterization of coordinated protein alterations produced by heteroplasmy in heart, liver and muscle using the SBT model. *Zqa* values are standardized log2-ratio averages of proteins from heteroplasmic mice relative to controls. Proteins are grouped into representative functional categories with a statistically significant change. The upper panel presents proteins in the OXPHOS category that are decreased in the heteroplasmic heart and that contain heteroplasmy-induced oxidative modifications.

3.2. Heteroplasmy mainly produces oxidative modifications of OXPHOS proteins in heart

To study the impact of heteroplasmy on the modified peptidome, we first compared the ΔMass distribution of peptides searched with CometPTM and identified with SHIFTS in the three tissues. The distribution of the most abundant peaks was very similar in liver, heart and muscle (Figure 17). In addition to the most intense peak corresponding to unmodified peptides, we found peaks corresponding to oxidized and deamidated peptides, and dioxidized peptides were also frequent. Other frequent modifications were artifacts produced by TMT reagents and iodoacetamide treatments, sodium adducts and missed cleavages. Some tissue-specific modifications were also detected, mostly in the region from 50 to 100 Da. Among these, phosphorylated peptides were more abundant in muscle and 90% of these peptides belonged to proteins implicated in muscle contractility, consistent with the regulatory role of phosphorylation in the sarcomere. Within each Δmass peak, we analyzed the distribution of the modified amino acid residues, which were assigned to frequent modifications using the unbiased approach followed in Figure 14B. Confirming the results described above, amino acids known to harbour specific modifications were mostly located correctly (Figure 18). Interestingly, the relative proportion of modified sites was remarkably preserved across the three tissues, with the exception of tyrosine iodination, which was prominent in heart but less frequent in liver and skeletal muscle (Figure 18). This analysis also detected the increased frequency of phosphorylation in muscle (Figure 18). We estimated that about one quarter of the total peptidome in the 3 tissues had known posttranslational modifications, mostly oxidation and deamidation; less than 10% of modifications were chemical artifacts (Figure 19A).

Second, we analysed the quantitative data with the integrative statistical algorithm (Figure 10A), to determine the pattern of modifications that are specifically affected by heteroplasmy in each tissue. Heteroplasmy induced marked tissue-specific modifications, most of them concentrated in heart and barely detectable in skeletal muscle (Figure 19B and Appendix Table 3). The vast majority of increased modifications in heart were oxidative, affecting mostly Tyr, Trp, Pro and Phe and to a lesser extent Asn, Cys and Asp (Figure 19C). Heteroplasmy also induced phosphorylation and methylation in the heart. The modification pattern was similar in liver, but the effect of heteroplasmy in this tissue was significantly less pronounced.

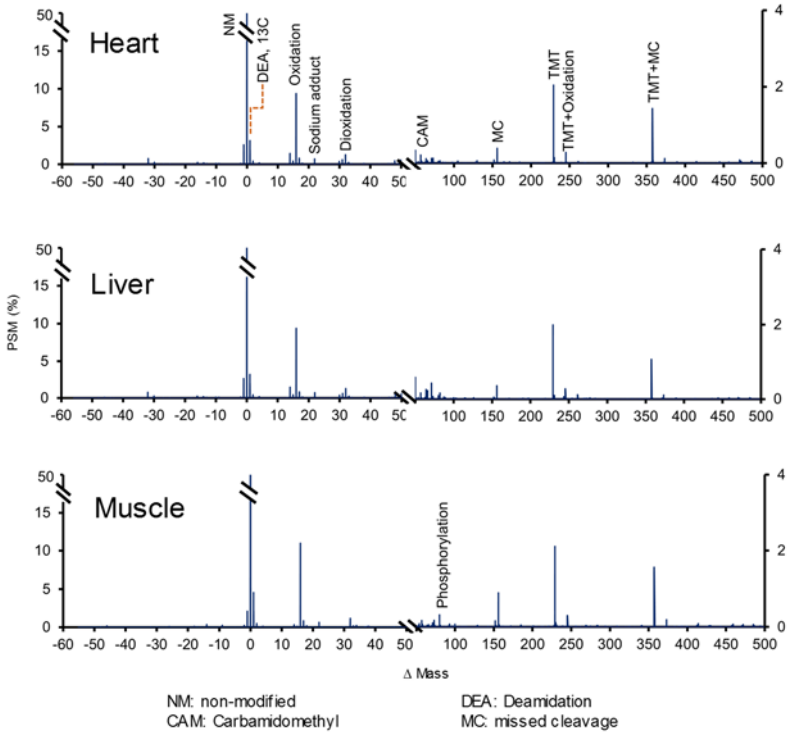


Figure 17: ΔMass distribution of modified peptides identified in the 3 mouse tissues. Assignations of the most frequent ΔMass peaks are indicated.

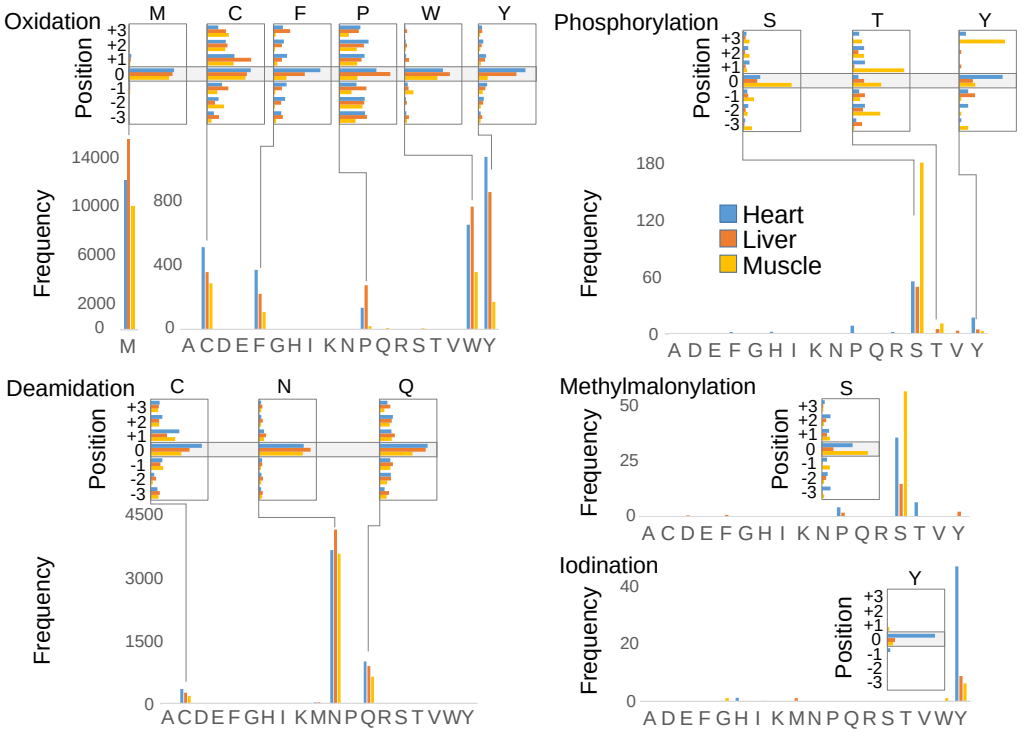


Figure 18: Distribution of amino acids containing the most frequent modifications. The horizontal and vertical bar graphs have the same meaning as in Figure 13B

Functional enrichment analysis revealed that heteroplasmy-induced modifications in mice heart were mostly located in mitochondrial proteins, particularly those located in the inner mitochondrial membrane; moreover, most of the modified proteins were OXPHOS proteins, which were enriched with multiple oxidative modifications (Figure 20A). TCA cycle proteins and those related to the oxidoreductase complex were enriched in several modifications. Deamidation affected several mitochondria-related categories, and phosphorylation affected proteins related to contractility. Heteroplasmy-induced modifications affected considerably fewer categories in liver (Figure 20B) and were not significantly enriched in muscle (Figure 20C), where only a few categories were decreased, containing low numbers of proteins. Deamidated peptides were found in proteins related to mitochondria, NAD binding and extracellular exosomes in both heart and liver, suggesting that these modifications do not contribute to the tissue-specific phenotypic differences between heteroplasmic and control mice.

Interestingly, heteroplasmy decreased phosphorylations in proteins related to chaperone activity in the heart but had the opposite effect in liver (Figure 20A and B). This category contains the 2 HSP90 isoforms (HSP90AA1 and HSP90AB1), which harbour PTM sites involved in the regulation of the chaperone activity (Mollapour and Neckers, 2012). These isoforms are known to be phosphorylated at Ser263 and Ser255, respectively (Wang *et al.*, 2010; Hu *et al.*, 2015), within the CK2 phosphorylation motif (S-X-X-E), which is crucial for protein activity. In particular, phosphorylation at Ser255 is required for the activation of MAPK/ERK pathway. Comet-PTM and SHIFTS identified these 2 phosphorylated sites, and heteroplasmy decreased their levels in heart while increasing them in liver (Figure 20D).

These findings suggest that heteroplasmy differentially affects chaperone activity and support the roles in cardiac heteroplasmy we have proposed for the endoplasmatic reticulum and the mitochondrial unfolded stress response (unpublished data).

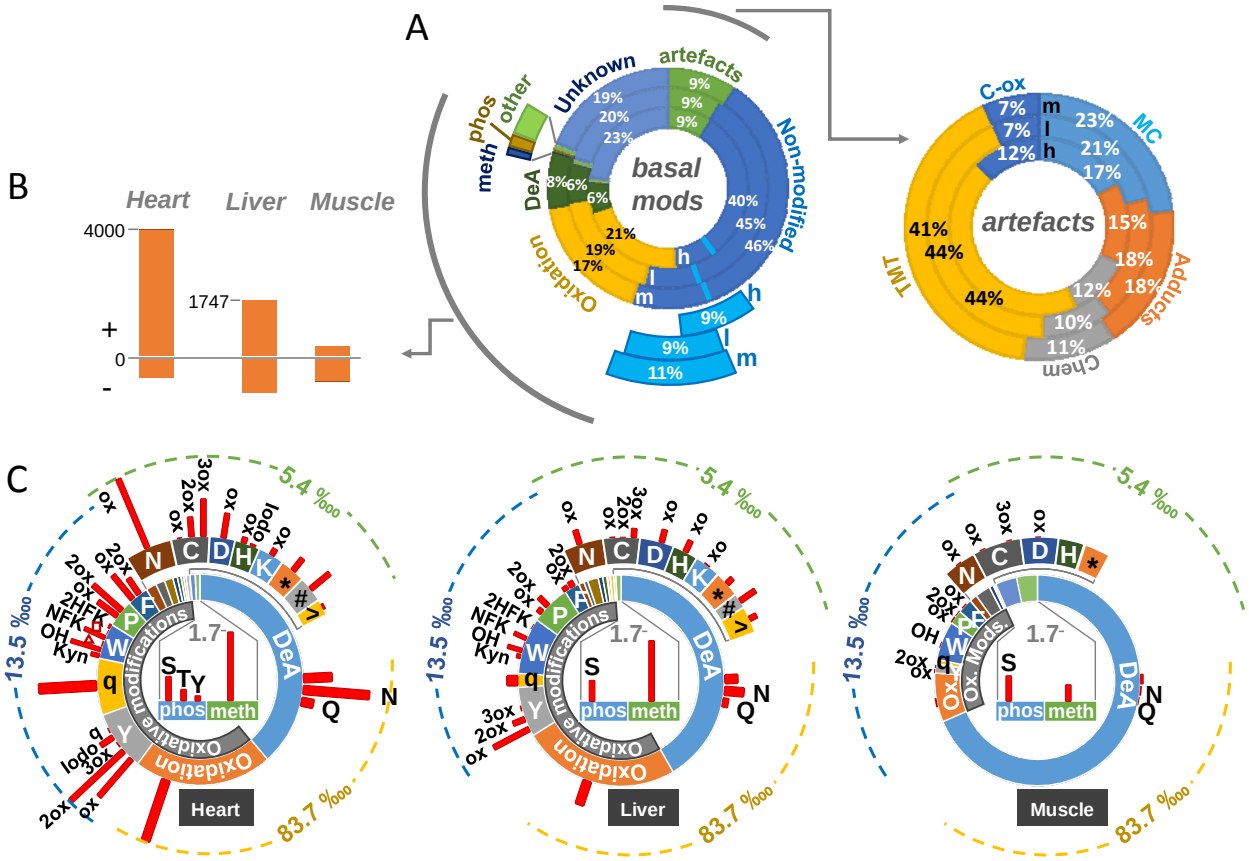


Figure 19: (A) Complete map of the basal peptidome, expressed as a percentage of the total number of peptides identified in each tissue (circular bar graphs: h, heart; l, liver; m, muscle). For simplicity, peptides oxidized on Met (light blue) are included in the group of nonmodified peptides. The plot to the right shows proportions of artifacts produced by sample manipulation. (B) Number of PTM-containing peptides significantly increased (+) or decreased (-) in heteroplasmic mouse tissues (n = 4) relative to controls (n = 3). (C) Statistically significant PTM increases in heteroplasmic mouse tissues according to the type of modification and the modified residue. The circular inner bars show the peptide proportion for each modification in the basal state, and the radial red bars represent the proportion of peptides of each kind (in parts per 10,000) that are increased in heteroplasmic tissues. For the sake of clarity, three different scales are used depending on the type of modification. DeA, deamidation; meth, methylation; phos, phosphorylation; MC, missed cleavages; Chem, chemical derivatives produced by sample preparation; TMT, extra addition of the isobaric labelling reagent; Adducts, Na, K, and ammonia adducts. Uppercase letters indicate the single-letter amino acid code. ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; red triangle, Trp to oxolactone; red square, Trp to quinone; q, quinone, Tyr to quinone; Iodo, iodination; asterisk, Met to homocysteic acid; hashtag, Met to homoserine.

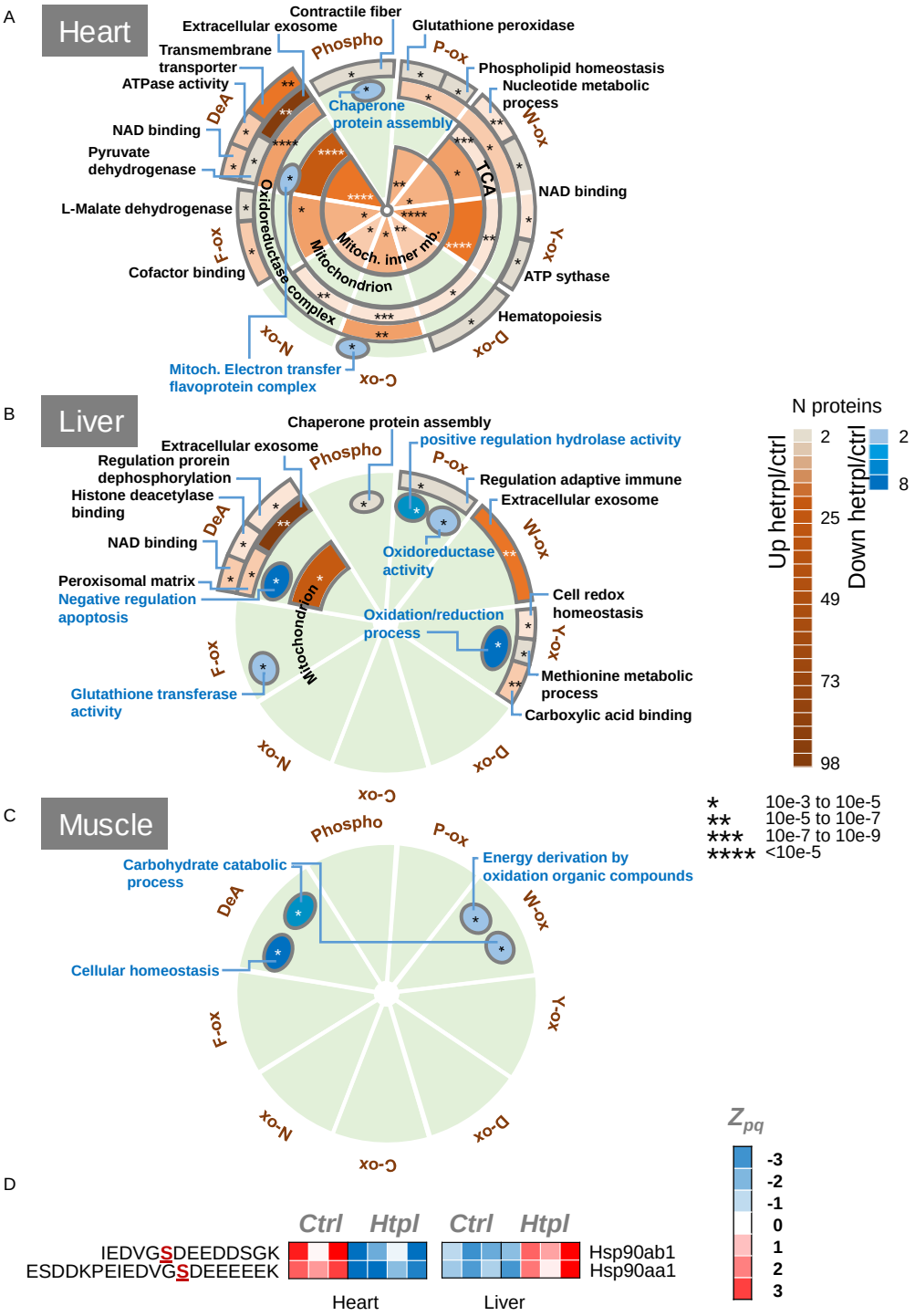


Figure 20: Functional distribution of proteins posttranslationally modified by heteroplasmy. (A-C) Functional enrichment analysis of the proteins containing PTMs that are significantly increased (brown scale) or decreased (blue scale) in different tissues of heteroplasmic mice in relation to control mice. The analysis was performed separately per different kinds of modification (slices). Color intensity represents the number of proteins in each category. Asterisks indicate statistical significance of category enrichment relative to the list of proteins identified in each tissue. Inner circular rings are used for categories at the cellular component level, whereas outer rings correspond to biological processes. (D) Heat map of the *Zpq* values for 2 modified proteotypic peptides from the 2 HSP90 isoforms (HSP90AB1 and HSP90AA1) for all control and heteroplasmic replicates in the heart and in liver. These 2 isoforms are included in the "chaperone protein assembly" functional category displayed in A and B.

Heteroplasmy also altered phosphorylation levels at sites known to regulate protein activity in a tissue-specific manner. Phosphorylation at Ser42 from PTRF, which regulated transcriptional activity in response to metabolic challenges (Huttlin *et al.*, 2010; Liu and Pilch, 2016), was decreased in heart but was increased in muscle, in agreement with the different metabolic status of these tissues (Appendix Table 3). Similarly Ser133 from MYG, which regulates contractility (Lundby *et al.*, 2013), was more phosphorylated in heart, in line with the coordinated decrease of proteins related to muscle function (Figure 16), suggesting alterations in the contractile function of the heart in heteroplasmic animals. In the liver, the active site Ser117 from PGM1, which regulates glycogen biosynthesis (Huttlin *et al.*, 2010; Lee *et al.*, 2014), was less phosphorylated, suggesting a lower fuel storage in this tissue (Appendix Table 3). These results reinforce the notion that the impact of heteroplasmy is tissue-specific.

The rest of phosphorylated sites affected by heteroplasmy had unknown regulatory roles, but it should be noted that all of them have been described previously, supporting the accuracy of the algorithms used to detect them. They include Ser181 in PGRMC1(Jin *et al.*, 2004), Ser442 in SRCA (Dai *et al.*, 2007), Ser475 in JPH1(Rigbolt *et al.*, 2011), and Ser298 and Ser304 in BIN1 (Ballif *et al.*, 2004; Huttlin *et al.*, 2010).

To compare the alterations in the modified peptidome with the changes observed at the protein level, we mapped the proteins harbouring post-translational modifications affected by heteroplasmy onto the protein distribution plots depicting coordinated protein responses (red and blue dots in Figure 16). Most increased modifications mapped onto the OXPHOS protein category in heart, which was decreased, whereas proteins belonging to other categories were mostly unaffected. This finding suggests that the oxidative damage produced by heteroplasmy in heart OXPHOS proteins induced their degradation, and is consistent with the increase in oxidative stress we have observed in heart slices, embryonic fibroblasts and adult fibroblasts of heteroplasmic mice (Unpublished data).

Most heteroplasmy-induced oxidative modifications in OXPHOS proteins occurred in the four respiratory protein complexes, except for Trp oxidations, which were mostly concentrated in C-IV (Figure 21,22). The modifications were clustered in groups of proteins located together in the structure of the complexes, and predominantly affected proteins orientated toward the mitochondrial matrix, where proteins are more exposed to mitochondrial ROS. In further analysis, we used Consurf to calculate evolutionary

conservation profiles for the affected residues (Ashkenazy *et al.*, 2016) and mapped the modifications onto the known 3D structures of the complexes, when available, or onto their predicted models (Meier and Söding, 2015). We found that 70% of the modified residues were conserved or interacted with conserved regions (Appendix Table 4). Moreover, most of the modifications were located on the protein surface, and those that were buried tended to have high conservation scores, suggesting they are important for structural stability or folding (Appendix Table 4). These findings suggest that these oxidative modifications compromise OXPHOS protein function. This view is supported by the decreased ATP synthesis capacity we have detected in heteroplasmic heart.

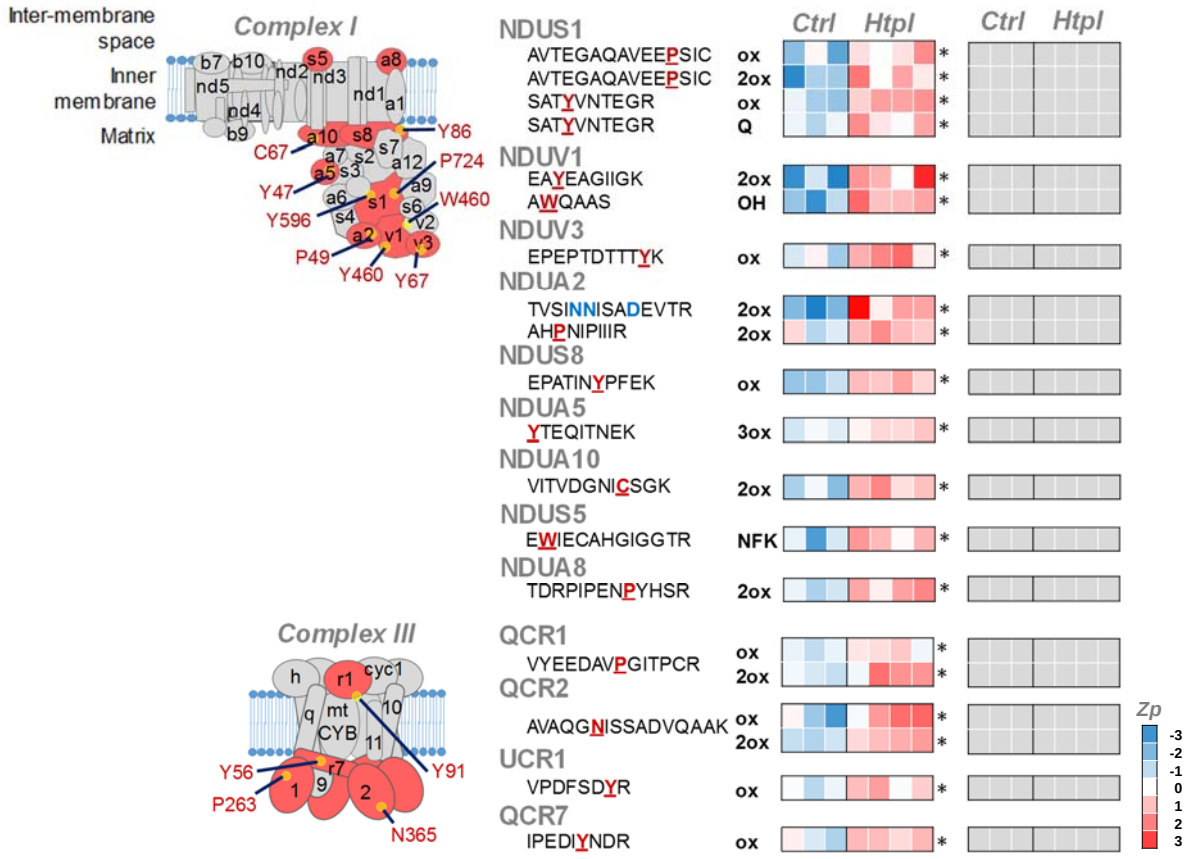


Figure 21: PTMs Induced by Heteroplasmy in OXPHOS Proteins. Scheme of heteroplasmy-induced PTMs in proteins of complex I (upper), and complex III (lower) from the OXPHOS system.

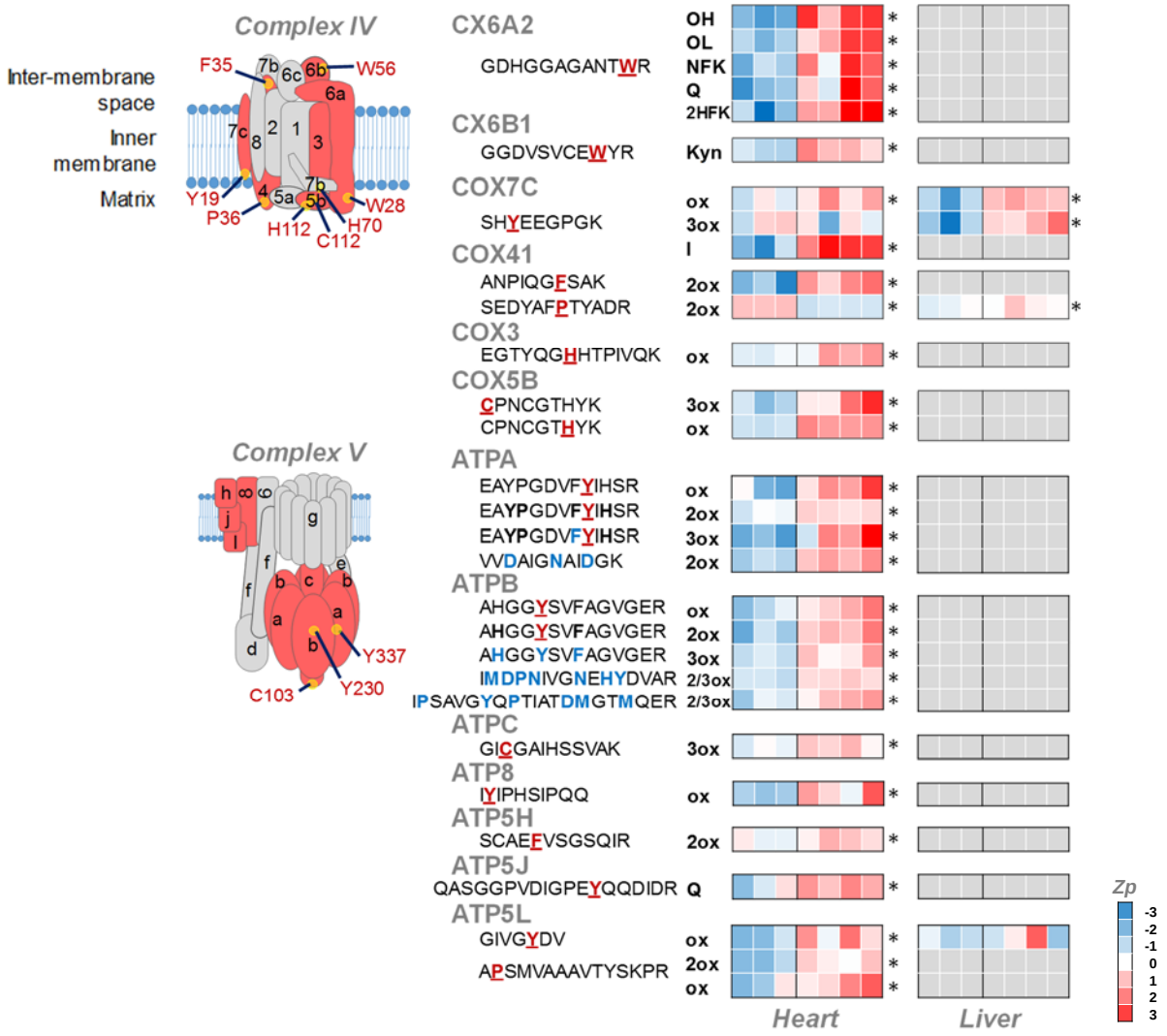


Figure 22: PTMs Induced by Heteroplasmy in OXPHOS Proteins. Scheme of heteroplasmy-induced PTMs in proteins of complex IV (upper) and complex V (lower) from the OXPHOS system. PTMcontaining proteins are highlighted in red, with the spatial position of the modified residue indicated. The heat maps show statistically significant changes in abundance (*Zpq)* of the indicated peptides in heart and liver from control mice (Ctrl) and heteroplastic mice (Htpl). Grey squares indicate that the modified peptide was not detected in the tissue sample. Modified residues that are unambiguously assigned are underlined bold red letters; tentatively assigned sites are in bold blue letters. Asterisks mark statistically significant changes between control and heteroplasmic samples according to Student's t test ($p < 0.05$). The modification type is indicated in the column to the right of the peptide sequences: ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; OL, Trp to oxolactone; Q, Trp to quinone; I, iodination.

Discussion

1. Hypothesis- driven proteomics: it's about time we move forward

MS has traditionally been used to identify and map modifications on proteins. Historically, PTM identification was one of the first applications of MS in protein research (Stenflo *et al.*, 1974). Even though PTM identification had a promising start in early MS development, the deeper investigation of PTMs and the high-throughput PTM analysis of have started only in the last decade. Although with the current state-of-art MS equipment near complete proteome analysis is within grasp, this is not the case with PTMs. The identification of PTMs in a comprehensive way is more challenging from both technological and conceptual points of view. There are several principal differences between proteome and PTM analysis. For instance, in a typical proteome measurement, a protein is characterized by several peptides, making identification and quantification of the proteome more robust, whereas in an experiment focusing on PTM measurements, each peptide representing a modification site of interest needs to stand on its own. Beside, modified peptides can be in lower abundance than their non-modified counterpart, creating a necessity for PTM enrichment that are often limited to just some kinds of PTMs.

Comprehensive analysis of PTMs in mass-spectrometry data obtained through shotgun approaches has historically presented a bioinformatics challenge because conventional (or closed) database searches can only identify predefined subsets of modifications. For instance, most efforts are dedicated to the most prominent and known modifications, such as phosphorylation, ubiquitination, glycosylation, and acetylation. However, several hundreds of modifications are known, and considering chemically induced modifications, the list grows drastically. Conventional PTM identification requires prior assumption of the kind of PTMs present in a given sample. Possible modifications and modified amino acids must be specified in the search engines (usually a maximum of 6 to 8 modifications). However, due to the continuous improvement in MS accuracy, improved search algorithms and PTM knowledgebase, tens of modifications are studied and used in the search engines in a single experiment, making the analysis very challenging and resource demanding, in some cases resulting in months of search time. Despite the much-debated cons of this kind of targeted analysis, it is conceptually straightforward and has therefore become an important part in the routine proteomic experiments.

Moreover, in experiments aiming at PTM analysis, it is very important that there is sufficient information in the tandem MS spectra, which can result in confident peptide identification and modified site localization. These limitations hinder truly hypothesis-free analysis of modifications, to the point that the proportion of chemical and post-translational protein modifications in biological systems has remained unknown.

To overcome above-mentioned issues, the proteomics community needs to develop novel algorithms and tools to make PTM analysis faster in similar ways as the genomics community has done with next-generation sequencing. Fortunately, in recent days new trends and developments are very encouraging, as is the deep phospho analysis performed in a single day measurement time (Bai *et al.*, 2017; Hogrebe *et al.*, 2018).

2. Hypothesis-free proteomics: a potential replacement of conventional database search approaches

The recently published open database search strategy (Chick *et al.*, 2015) provided a starting point for resolving this problem, allowing a truly unbiased high-throughput identification of modified peptides without previous knowledge on the nature of these modifications. Later, an improved fragment-ion indexing method (MSFragger) demonstrated that open searches can be performed on a timescale similar to or shorter than that required for closed searches (Kong *et al.*, 2017). The open search strategy has the additional advantage that modified and nonmodified peptide forms are identified simultaneously (Kong *et al.*, 2017). These strategies, however, has a major drawback as they rely on the chance that the modification leaves enough unaffected fragment ion to be matched by search engine, or in other terms; these algorithm identifies modified peptides from MS/MS spectra, but only with the non-modified fragments. These OS strategies therefore diminishes the score assigned to a modified peptide and only half of the modified peptides detectable by conventional searches (CS) are identified. In addition, OS methods also fails to pinpoint or locate the modified amino acid in a peptide. Considering all current limitations in the current state, OS approaches are not comprehensive and provide limited information on the nature of modified residues.

The novel computational approach presented in this thesis enables the open search strategy to match the performance of closed searches, identifying the vast majority of the MS-detectable peptide modifications. Although Comet-PTM is built on the framework of a well-known database-search engine, it is important to note that the underlying concept in it can be

extrapolated to any other searching engine, opening the way to near-complete analysis of the protein-modification landscape in biological systems using tools already used by the proteomics. The fact that Comet-PTM takes into account the mass shift in fragments, the identification scores are improved and as result, Comet-PTM not only outperforms the classical OS approaches but also produces a performance equal to targeted ssearch (CS). In several instances, Comet-PTM also corrects PTMs wrongly assigned by CS. For instance, oxidation was assigned to different amino acids such as Trp, Pro and Tyr that were forcibly assigned to Met in CS. This finding shows that conventional PTM searches using a variable modification on selected amino acids can force false assignation of the modified residue. Taken together, these data show that the CometPTM OS engine efficiently resolves the modification mass-shift in the fragmentation series, doubling the number of modified peptides identified by conventional OS and matching the identification performance of targeted CS.

Localizing the modifications to specific residues is often considered a less reliable process than identifying peptides, partly because there is frequently insufficient information to determine the exact modified residue. For this reason, there is a clear lack of algorithms, which not only identify all possible PTMs, but also localize the modified amino acid in single-pass searches. Usually, site localization algorithms are used post identification and the majority of these algorithms are optimized for specific modifications; such is the case of PhosphoRS (Taus *et al.*, 2011) or Ascore (Taus *et al.*, 2011). Comet-PTM resolves these issues and provides information on the most probable site containing the modification with an estimated accuracy of ~85%. It should be noted that, in its current state of development, the method might fail if the peptide contains more than one modification. However, the impact of this limitation on the general performance of the algorithm seems small, since, according to our data the fraction of multiply modified peptides that were not identified was almost negligible.

In current researches of shotgun proteomics experiment the use of separate FDR for PTMs has been suggested and discussed thoroughly. The idea of a separate FDR or Local FDR comes in paly due to the basic difference in nature of spectra of modified and non-modified; that is the modified spectra can have a neutral loss or insufficient fragmentation. Due to their own specific features in spectra, modified peptides can have different score distribution compared to non-modified peptide. In addition, abundance difference in modified and non-modified peptide in a given sample adds another challenge in FDR estimation, as chances of matching to a correct and incorrect peptide sequence are different. In order to overcome the problem in PTM FDR,

we developed and introduced a conservative, three-layered approach to control the FDR of peptide identification. Our local FDR calculation is similar to that used by MSFragger (Kong *et al.*, 2017), with the only difference that we used the decoy/target competition strategy to estimate FDR. However, Kong et al. modelled local score distributions, in the same 1-Da bins as a mixture of correct and incorrect identifications and calculated the probability of correct identification by the Bayes rule. As recognized by these authors, their approach did not take into account the p.p.m. levels of accuracy produced by high-resolution instruments (Kong *et al.*, 2017), which we have considered in our model by applying an additional peak FDR estimation. In addition, we found it necessary to introduce a third global FDR filter to avoid identification of low-scoring peptides in local bins or in peaks that, as in the case of oxidation, concentrate high numbers of target PSMs.

The new algorithm allowed us to produce peptide maps containing realistic proportions of unmodified peptides, posttranslational modifications, and chemical artifacts in mouse tissues and of amino acid residues subjected to each type of modification. Many of the modifications are introduced during sample preparation, including those induced by exposure to Cys-alkylating and amine-directed isobaric labelling reagents; while these agents are thought to be highly specific, in practice they produce secondary reactions with other nucleophiles.

We show that the basal distribution of modifications and modified amino acids is remarkably similar across the 3 mouse tissues studied, indicating that the algorithm generates reproducible results and that most peptide modifications are of an oxidative nature.

In the direction of revelling the importance of PTMs, the statistical approach presented here enables quantitative analysis of the modified peptidome within the same framework used to analyse the unmodified proteome; this approach thus allows coherent interpretation of all the results and opens the way to integrated systems-biology analysis. We demonstrate the performance of the new algorithm by simultaneously characterizing, in three tissues of a mouse model, the quantitative impact of heteroplasmy on the proteome and on the modified peptidome. To our knowledge, this is the first study to apply open-search-based approaches together with quantitative mass spectrometry, allowing comprehensive and unbiased characterization of alterations in the protein modification landscape produced in a biological system.

3. Hypothesis-free PTM analysis: from algorithm development to application

The ability to perform a truly hypothesis-free high-throughput analysis of modifications can contribute to a deeper understanding of the relation between protein chemistry and function. However, the most important biological insight obtainable with shotgun proteomics is how these modifications are altered during normal physiological processes or because of disease. Detecting these specific alterations allows concentrating the study on biologically meaningful modifications, obviating the analysis of the abundant artifacts produced by sample preparation. Understating disease mechanisms has been mainly done with the help of either animal models or cell culture. However, more and more different omics analysis including proteomics are contributing towards clinical sample analysis. If the precision and personalized medicine is the future of medical science, PTMs might open a completely new path towards biomarker identification. Although, multiple PTMs are reported as potential biomarker of disease, only a few are routinely used for diagnostic purpose (Silva, 2015). Considering the complexity of a disease, in many cases a single molecule or a single PTMs or modification site may be insufficient as specific biomarker for diagnosis. Recent studies suggests, that combining different omics (genomics, lipidomics, metabolomics) with an extensive PTM map and identification would be to our advantage for identifying specific disease patterns and signatures (Huang et al., 2017).

Nowadays an increasing amount of diseases is being linked to mitochondrial DNA (Ferries *et al.*, 2017). Several medical technologies, mainly in the infertility treatments, have emerged aiming to prevent the transmission of these mtDNA-linked diseases (Tachibana *et al.*, 2009; Craven *et al.*, 2010; Woods and Tilly, 2015). However, the main "side effect" of these techniques is the generation of heteroplasmy, which is the presence of more than one variant of mtDNA co-existing in the same cytoplasm. Interestingly, some tissues (liver, blood cells) are able to slowly resolve the heteroplasmy, while others (heart, skeletal muscle, brain) cannot (Sharpley *et al.*, 2012; Burgstaller *et al.*, 2014). Some authors have raised concerns about the phenotypic impact on the offspring of some kinds of heteroplasmy (Battersby, Redpath and Shoubridge, 2005; Sharpley *et al.*, 2012). In addition, recent results from the Functional Genetics of the Oxidative Phosphorylation System laboratory (CNIC) led by Juan Antonio Enriquez have evidenced that some kinds of wild type heteroplasmy can cause severe pathology upon aging, consisting in heart, lung and skeletal muscle wasting. Consequently, heteroplasmic mice present premature ageing-like symptoms and a higher frequency of early

death (manuscript in preparation). The molecular cause of these negative effects of heteroplasmy in the different tissues of the offspring remains unknown. Thus, we selected this biological model to test the performance of our newly developed tools and the same time, to shed some light on the possible molecular mechanisms underlying the pathological effects of heteroplasmy, when they are being set up at an early age (12-week-old animals). Thus, we studied the protein levels and performed a systems biology analysis in a tissue able to resolve heteroplasmy during aging like the liver, and in two tissues where heteroplasmy remains stable with age, like the heart and skeletal muscle.

First, we observed a generalized decrease in OXPHOS proteins in heteroplasmic hearts, but not in liver or skeletal muscle. Consistently, our collaborators we have reported a decreased ATP synthesis and an abnormal increase in phosphocreatine/ATP ratio in the heart and an increase in plasma creatine kinase in this animal model (manuscript in preparation). Thus, we provide further molecular evidence that heteroplasmic cardiac mitochondria have a compromised ability to supply energy to the sarcomere. In addition, the increased levels of vesicle trafficking proteins involved in the insulin-stimulated glucose uptake, agree with the higher glucose uptake our collaborators detected in heteroplasmic hearts, and point to a shift toward glycolytic metabolism in order to compensate the reduced phosphocreatine supply from the OXPHOS.

Next, we hypothesized that tissue-specific PTMs could be a possible explanation for the reduction in the levels of OXPHOS proteins in heteroplasmic hearts. To answer this question, we first took advantage of the ability given by our newly developed tools to estimate for the first time the real proportions of the non-modified and modified peptide forms in a proteome, and studied the basal "modificome" maps of the three tissues. As expected, the relative proportions of artefactual modifications was the same among the three tissues, but to our surprize, the physiological PTMs were quite similar too.

However, the changing PTMs brought about by heteroplasmy were quite different in the three tissues not only in its nature, but also in its extent. The heart was the most "modified" tissue in heteroplasmic animals, being most changing PTMs of oxidative nature, mainly on Tyr, Trp and Pro, and affecting mostly OXPHOS proteins. Although most changing modifications in liver were also oxidative, they were much fewer compared to the heart and were not concentrated on specific categories. The skeletal muscle was the least affected tissue, containing mainly increased deamidations, scattered on just a few proteins.

Thus, the application of our tools reveals that the major hallmark of mitochondrial heteroplasmy in 12-week-old animals is oxidative damage to OXPHOS proteins in the heart. This is mainly reflected in the increased levels of oxidative modifications and in the decreased levels of OXPHOS proteins. Our data support the notion that heteroplasmy involving different nonpathological mtDNA variants affects the performance of the OXPHOS system in the heart; resulting in compromised mitochondrial ROS handling that triggers an oxidative stress response and impairs the ability to supply energy to the sarcomere. Heteroplasmy is known to be related to ROS production (Hamalainen et al., 2013; Hämäläinen et al., 2015) and also affects the levels of contractile proteins in muscle. These alterations do not occur in liver, which we have shown to be able to resolve heteroplasmy (unpublished data). Our data thus provide a molecular mechanism that explains the functional findings we have observed in the heteroplasmic heart, which include a diminished ability of mitochondria to produce ATP, increased glucose uptake, an exacerbated phosphocreatine/ATP ratio, and mitochondrial disruption and structural alterations.

4. Future perspective

Taking into account, the algorithms and tools developed for PTMs identification and quantification, to best of our knowledge, for the first time, we have formalized a way to automate the whole process to analyse the modified peptidome in a true hypothesis-free manner. Further, applying these developments to a biological model has greatly improved our understanding of PTMs in the context of not only technical but also chemical level. At the same times, we have also realized the potential challenges and problem, which we will have to deal within the near future.

4.1. Decoding the Δmass combinations and aim for a faster search time

With Comet-PTM, we have been able to localize the modification on their site with a very high accuracy, as Comet-PTM uses Δmass on fragments and then calculates the Xcorr for retrieving the modified peptide. However, Comet-PTM fails in very limited number of cases where Δmass is a combination of two different modification, resulting in either wrong identification or lower Xcorr. To a certain extent, this problem is taken care of post identification using PtmSticker: in which Δmass combination is broken into known modification masses, still leaving out the possibility of assigning the modification to it's corresponding site. In present, we are working on an integration of an algorithm to Comet-PTM, which can decipher the Δmass pre handed and then locate the modification site.

On the other hand, due to considering the Δmass on fragments and Xcorr calculation on every amino acid makes Comet-PTM very computationally expensive. As a result, a high performance computer cluster is needed. To resolve the issue, we are working on a database-indexing algorithm, which will facilitate a faster candidate peptide retrieval.

4.2. Re-ranking of PSMs: Xcorr to corrected Xcorr

Like any other search engine, Comet-PTM provides multiple candidates for a scan, which are sorted based on their score (Xcorr in Comet-PTM). In brief, Xcorr is calculated based on the number of fragments matched between theoretical and experimental spectra. As a result, longer peptide scores better than shorter peptides, increasing the probability of a false identification just by chance. We are currently working on an idea where corrected Xcorr (Keller *et al.*, 2002) will be calculated within Comet-PTM and PSMs would be ranked accordingly. This improvement will help us to retrieve PSMs from far negative side to FDR controlled region (FDR control in SHIFTS).

Similarly, we are working on a potential idea where; instead of the Δmass (obtained from the difference between experimental and theoretical mass), the theoretical mass of the modification can be used in the fragments to calculate the identification scores. The reason behind the idea is that in some cases, either due to the instrument error, or to the low abundance of the modified precursor, the Δmass is introduced in the fragments with a slight deviation (has a slight error in Δmass), mismatching some of the fragments. Providing a curated list of theoretical Unimod delta masses would result in slightly better scores compared to the experimental Δmass and would recover some identifications, which are likely to be left out in FDR due to low score.

In the same direction, I am working on PtmSticker to include an error model, which will take into account the peak abundance to annotate the Δmass distribution automatically (using Monte Carlo simulation). In this way, we will avoid using a fixed error to assign the peaks to PTMs and achieve a reliable annotation.

4.3. Rediscovering the plasma proteome: a search for biomarker beyond proteins

Currently the suit of developed algorithms and tools is being applied on a plasma proteome in a follow-up study to find biomarkers for subclinical atherosclerosis. Although many

efforts have been made to find systemic biomarkers, none of the candidates has provided clinical diagnostic relevance yet. On the other hand, evidences are accumulating that the modified lipoproteins could be involved in the progression of the disease (Nicholls and Hazen, 2009; Delporte *et al.*, 2013). Such evidences suggest that post-translational (PTMs) and structural modifications could potentially be used as biomarkers. Using these new developments for first time a quantitative map of human plasma PTMs is being created which will open a window to the pathogenic mechanisms underlying atherosclerosis (PhD thesis, Ileana Beatriz González Valdés).

Conclusions

1. The CometPTM pipeline enables a true hypothesis-free and unbiased identification of PTMs.

2. Comet-PTM opens a way to near complete identification of all peptide modifications detectable by MS.

3. Comet-PTM provides information on the amino acid affected by the PTM with an accuracy compatible with its use in high-throughput proteomics experiments.

4. The new statistical method allows automated quantification of modified peptides, independent of protein change in abundance, and simultaneous analysis of the proteome and of the modified peptidome.

5. The new bioinformatics tools generate for the first time the complete landscape of post-translational modifications containing realistic proportions of unmodified and modified peptides and of chemical artifacts.

6. Mitochondrial heteroplasmy in mice affects mainly the heart, inducing oxidative damage to proteins of the oxidative phosphorylation system.

References

and, E. S. B. and Fales*, H. M. (2001) 'Overalkylation of a Protein Digest with Iodoacetamide'. American Chemical Society . doi: 10.1021/AC0103423.

Ashkenazy, H. *et al.* (2016) 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.', *Nucleic acids research*. Oxford University Press, 44(W1), pp. W344-50. doi: 10.1093/nar/gkw408.

Bai, B. *et al.* (2017) 'Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry.', *Methods in enzymology*. NIH Public Access, 585, pp. 377–395. doi: 10.1016/bs.mie.2016.10.007.

Ballif, B. A. *et al.* (2004) 'Phosphoproteomic Analysis of the Developing Mouse Brain', *Molecular & Cellular Proteomics*, 3(11), pp. 1093–1101. doi: 10.1074/mcp.M400085-MCP200.

Battersby, B. J., Redpath, M. E. and Shoubridge, E. A. (2005) 'Mitochondrial DNA segregation in hematopoietic lineages does not depend on MHC presentation of mitochondrially encoded peptides', *Human Molecular Genetics*, 14(17), pp. 2587–2594. doi: 10.1093/hmg/ddi293.

Batth, T. S., Francavilla, C. and Olsen, J. V. (2014) 'Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics', *Journal of Proteome Research*. American Chemical Society, 13(12), pp. 6176–6186. doi: 10.1021/pr500893m.

Bjornson, R. D. *et al.* (2008) 'X!!Tandem, an Improved Method for Running X!Tandem in Parallel on Collections of Commodity Computers', *Journal of Proteome Research*, 7(1), pp. 293–299. doi: 10.1021/pr0701198.

Bonzon-Kulichenko, E. *et al.* (2015) 'Revisiting Peptide Identification by High-Accuracy Mass Spectrometry: Problems Associated with the Use of Narrow Mass Precursor Windows', *Journal of Proteome Research*. American Chemical Society, 14(2), pp. 700–710. doi: 10.1021/pr5007284.

Broncel, M. *et al.* (2010) 'How Post-Translational Modifications Influence Amyloid Formation: A Systematic Study of Phosphorylation and Glycosylation in Model Peptides', *Chemistry - A European Journal*. Wiley-Blackwell, 16(26), pp. 7881–7888. doi: 10.1002/chem.200902452.

Burgstaller, J. P. *et al.* (2014) 'MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage.', *Cell reports*. Europe PMC Funders, 7(6), pp. 2031–2041. doi: 10.1016/j.celrep.2014.05.020.

Carapito, C. *et al.* (2009) 'Systematic LC-MS Analysis of Labile Post-Translational Modifications in Complex Mixtures', *Journal of Proteome Research*, 8(5), pp. 2608–2614. doi: 10.1021/pr800871n.

Chalkley, R. J. *et al.* (2008) 'In-depth Analysis of Tandem Mass Spectrometry Data from

Disparate Instrument Types', *Molecular & Cellular Proteomics*, 7(12), pp. 2386–2398. doi: 10.1074/mcp.M800021-MCP200.

Chammas, R. *et al.* (1999) 'De-N-acetyl-gangliosides in humans: unusual subcellular distribution of a novel tumor antigen.', *Cancer research*, 59(6), pp. 1337–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10096568 (Accessed: 8 May 2018).

Chee Sian Gan *et al.* (2007) 'Technical, Experimental, and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ)'. American Chemical Society. doi: 10.1021/PR060474I.

Chen, Y. *et al.* (2009) 'PTMap--a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(3), pp. 761–6. doi: 10.1073/pnas.0811739106.

Chi, A. *et al.* (2007) 'Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(7), pp. 2193–8. doi: 10.1073/pnas.0607084104.

Chick, J. M. *et al.* (2015) 'A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides.', *Nature biotechnology*. NIH Public Access, 33(7), pp. 743–9. doi: 10.1038/nbt.3267.

Choi, H., Fermin, D. and Nesvizhskii, A. I. (2008) 'Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics', *Molecular & Cellular Proteomics*, 7(12), pp. 2373–2385. doi: 10.1074/mcp.M800203-MCP200.

Choi, H. and Nesvizhskii, A. I. (2008) 'False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics', *Journal of Proteome Research*, 7(1), pp. 47–50. doi: 10.1021/pr700747q.

Clough, T. *et al.* (2009) 'Protein Quantification in Label-Free LC-MS Experiments', *Journal of Proteome Research*. American Chemical Society, 8(11), pp. 5275–5284. doi: 10.1021/pr900610q.

Cox, J. *et al.* (2011) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*. Nature Publishing Group, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Craven, L. *et al.* (2010) 'Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease', *Nature*, 465(7294), pp. 82–85. doi: 10.1038/nature08958.

Dai, J. *et al.* (2007) 'Protein Phosphorylation and Expression Profiling by Yin-Yang Multidimensional Liquid Chromatography (Yin-Yang MDLC) Mass Spectrometry', *Journal of Proteome Research*, 6(1), pp. 250–262. doi: 10.1021/pr0604155.

Daly, D. S. *et al.* (2008) 'Mixed-Effects Statistical Model for Comparative LC−MS Proteomics Studies †', *Journal of Proteome Research*. American Chemical Society, 7(3), pp. 1209–1217. doi: 10.1021/pr070441i.

Delporte, C. *et al.* (2013) 'Low-density lipoprotein modified by myeloperoxidase in inflammatory pathways and clinical studies.', *Mediators of inflammation*. Hindawi, 2013, p. 971579. doi: 10.1155/2013/971579.

Diament, B. J. and Noble, W. S. (2011) 'Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra', *J. Proteome Res*, 10, pp. 3871–3879. doi: 10.1021/pr101196n.

Duan, G. and Walther, D. (2015) 'The Roles of Post-translational Modifications in the Context of Protein Interaction Networks', *PLOS Computational Biology*. Edited by P. Radivojac. Public Library of Science, 11(2), p. e1004049. doi: 10.1371/journal.pcbi.1004049.

Elias, J. E. and Gygi, S. P. (2007) 'Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry', *Nature Methods*, 4(3), pp. 207–214. doi: 10.1038/nmeth1019.

Eng, J. K. *et al.* (2008) 'A Fast SEQUEST Cross Correlation Algorithm', *Journal of Proteome Research*. American Chemical Society, 7(10), pp. 4598–4602. doi: 10.1021/pr800420s.

Eng, J. K. *et al.* (2015) 'A Deeper Look into Comet—Implementation and Features', *Journal of The American Society for Mass Spectrometry*, 26(11), pp. 1865–1874. doi: 10.1007/s13361-015-1179-x.

Eng, J. K., Jahan, T. A. and Hoopmann, M. R. (2013a) 'Comet: An open-source MS/MS sequence database search tool', *PROTEOMICS*, 13(1), pp. 22–24. doi: 10.1002/pmic.201200439.

Eng, J. K., Jahan, T. A. and Hoopmann, M. R. (2013b) 'Comet: An open-source MS/MS sequence database search tool', *PROTEOMICS*. Wiley-Blackwell, 13(1), pp. 22–24. doi: 10.1002/pmic.201200439.

Eng, J. K., McCormack, A. L. and Yates, J. R. (1994) 'An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database', *Journal of the American Society for Mass Spectrometry*, 5(11), pp. 976–989. doi: 10.1016/1044-0305(94)80016-2.

Engholm-Keller, K. *et al.* (2012) 'TiSH — a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC', *Journal of Proteomics*, 75(18), pp. 5749–5761. doi: 10.1016/j.jprot.2012.08.007.

Ferries, S. *et al.* (2017) 'Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer', *Journal of Proteome Research*. American Chemical Society, 16(9), pp. 3448–3459. doi: 10.1021/acs.jproteome.7b00337.

Fu, Y. (2012) 'Bayesian false discovery rates for post-translational modification proteomics *', *Statistics and Its Interface*, 5, pp. 47–59. Available at:

http://www.amt.ac.cn/member/fuyan/pubs/SII-5-1-a05-Fu.pdf (Accessed: 19 April 2018).

Fu, Y. and Qian, X. (2014) 'Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 13(5), pp. 1359–68. doi: 10.1074/mcp.O113.030189.

Gajjala, P. R. *et al.* (2015) 'Emerging role of post-translational modifications in chronic kidney disease and cardiovascular disease', *Nephrology Dialysis Transplantation*, 30(11), pp. 1814–1824. doi: 10.1093/ndt/gfv048.

García-Marqués, F. *et al.* (2016) 'A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 15(5), pp. 1740–60. doi: 10.1074/mcp.M115.055905.

Griss, J. *et al.* (2016) 'Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets', *Nature Methods*. Nature Publishing Group, 13(8), pp. 651–656. doi: 10.1038/nmeth.3902.

Gupta, M. K. and Robbins, J. (2014) 'Post-translational control of cardiac hemodynamics through myosin binding protein C', *Pflügers Archiv - European Journal of Physiology*, 466(2), pp. 231–236. doi: 10.1007/s00424-013-1377-y.

Hegab, Z. *et al.* (2012) 'Role of advanced glycation end products in cardiovascular disease.', *World journal of cardiology*. Baishideng Publishing Group Inc, 4(4), pp. 90–102. doi: 10.4330/wjc.v4.i4.90.

Hogrebe, A. *et al.* (2018) 'Benchmarking common quantification strategies for large-scale phosphoproteomics', *Nature Communications*. Nature Publishing Group, 9(1), p. 1045. doi: 10.1038/s41467-018-03309-6.

Hsieh, E. J. *et al.* (2010) 'Comparison of database search strategies for high precursor mass accuracy MS/MS data.', *Journal of proteome research*. NIH Public Access, 9(2), pp. 1138–43. doi: 10.1021/pr900816a.

Hu, C.-W. *et al.* (2015) 'Temporal Phosphoproteome Dynamics Induced by an ATP Synthase Inhibitor Citreoviridin', *Molecular & Cellular Proteomics*, 14(12), pp. 3284–3298. doi: 10.1074/mcp.M115.051383.

Hultin-Rosenberg, L. *et al.* (2013) 'Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 12(7), pp. 2021–31. doi: 10.1074/mcp.M112.021592.

Huttlin, E. L. *et al.* (2010) 'A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression', *Cell*, 143(7), pp. 1174–1189. doi: 10.1016/j.cell.2010.12.001.

Jenuth, J. P. *et al.* (1996) 'Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA', *Nature Genetics*, 14(2), pp. 146–151. doi:

10.1038/ng1096-146.

Jeong, K., Kim, S. and Bandeira, N. (2012) 'False discovery rates in spectral identification.', *BMC bioinformatics*. BioMed Central, 13 Suppl 16(Suppl 16), p. S2. doi: 10.1186/1471-2105-13-S16-S2.

Jin, W.-H. *et al.* (2004) 'Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry', *Rapid Communications in Mass Spectrometry*, 18(18), pp. 2169–2176. doi: 10.1002/rcm.1604.

Junmin Peng, † *et al.* (2002) 'Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome'. American Chemical Society. doi: 10.1021/PR025556V.

Karpievitch, Y. *et al.* (2009) 'A statistical framework for protein quantitation in bottom-up MS-based proteomics', *Bioinformatics*, 25(16), pp. 2028–2034. doi: 10.1093/bioinformatics/btp362.

Keller, A. *et al.* (2002) 'Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.', *Analytical chemistry*, 74(20), pp. 5383–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12403597 (Accessed: 26 April 2018).

Khadjavi, A. *et al.* (2011) 'Evidence of Abnormal Tyrosine Phosphorylated Proteins in the Urine of Patients With Bladder Cancer: The Road Toward a New Diagnostic Tool?', *The Journal of Urology*, 185(5), pp. 1922–1929. doi: 10.1016/j.juro.2010.12.029.

Kim, J.-S. *et al.* (2013) 'In-Source Fragmentation and the Sources of Partially Tryptic Peptides in Shotgun Proteomics', *Journal of Proteome Research*, 12(2), pp. 910–916. doi: 10.1021/pr300955f.

Kim, S. and Pevzner, P. A. (2014) 'MS-GF+ makes progress towards a universal database search tool for proteomics', *Nature Communications*. Nature Publishing Group, 5, p. 5277. doi: 10.1038/ncomms6277.

Klammer, A. A., Park, C. Y. and Noble, W. S. (2009) 'Statistical calibration of the SEQUEST Xcorr function', *Journal of Proteome Research*, 8(4), pp. 2106–2113. doi: 10.1021/pr8011107.

Kong, A. T. *et al.* (2017) 'MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics', *Nature Methods*. Nature Publishing Group, 14(5), pp. 513–520. doi: 10.1038/nmeth.4256.

Lee, Y. *et al.* (2014) 'Compromised Catalysis and Potential Folding Defects in *in Vitro* Studies of Missense Mutants Associated with Hereditary Phosphoglucomutase 1 Deficiency', *Journal of Biological Chemistry*, 289(46), pp. 32010–32019. doi: 10.1074/jbc.M114.597914.

Leyfer, D. and Weng, Z. (2005) 'Genome-wide decoding of hierarchical modular structure of transcriptional regulation by cis-element and expression clustering', *Bioinformatics*, 21(Suppl 2), pp. ii197-ii203. doi: 10.1093/bioinformatics/bti1131.

Li, Q. *et al.* (2017) 'Global Post-Translational Modification Discovery.', *Journal of proteome research*. American Chemical Society, 16(4), pp. 1383–1390. doi:

10.1021/acs.jproteome.6b00034.

Lina P. Aristoteli, †, Mark P. Molloy, †, ‡ and and Mark S. Baker*, †, ‡ (2006) 'Evaluation of Endogenous Plasma Peptide Extraction Methods for Mass Spectrometric Biomarker Discovery'. American Chemical Society. doi: 10.1021/PR0602996.

Link, A. J. *et al.* (1999) 'Direct analysis of protein complexes using mass spectrometry', *Nature Biotechnology*. Nature Publishing Group, 17(7), pp. 676–682. doi: 10.1038/10890.

Liu, L. and Pilch, P. F. (2016) 'PTRF/Cavin-1 promotes efficient ribosomal RNA transcription in response to metabolic challenges', *eLife*. eLife Sciences Publications Limited, 5, p. e17508. doi: 10.7554/eLife.17508.

Lohrig, K., Sickmann, A. and Lewandrowski, U. (2011) 'Strong Cation Exchange Chromatography for Analysis of Sialylated Glycopeptides', in *Methods in molecular biology (Clifton, N.J.)*, pp. 299–308. doi: 10.1007/978-1-61779-148-2_20.

López-Carballo, G. *et al.* (2002) 'Activation of the Phosphatidylinositol 3-Kinase/Akt Signaling Pathway by Retinoic Acid Is Required for Neural Differentiation of SH-SY5Y Human Neuroblastoma Cells', *Journal of Biological Chemistry*, 277(28), pp. 25297–25304. doi: 10.1074/jbc.M201869200.

Lu, B. *et al.* (2007) 'Automatic Validation of Phosphopeptide Identifications from Tandem Mass Spectra', *Analytical Chemistry*, 79(4), pp. 1301–1310. doi: 10.1021/ac061334v.

Lundby, A. *et al.* (2013) 'In Vivo Phosphoproteomics Analysis Reveals the Cardiac Targets of -Adrenergic Receptor Signaling', *Science Signaling*, 6(278), pp. rs11-rs11. doi: 10.1126/scisignal.2003506.

Ma, C. W. M. and Lam, H. (2014) 'Hunting for Unexpected Post-Translational Modifications by Spectral Library Searching with Tier-Wise Scoring', *Journal of Proteome Research*. American Chemical Society, 13(5), pp. 2262–2271. doi: 10.1021/pr401006g.

Mann, M. and Jensen, O. N. (2003) 'Proteomic analysis of post-translational modifications', *Nature Biotechnology*. Nature Publishing Group, 21(3), pp. 255–261. doi: 10.1038/nbt0303-255.

Marshall Bern, *, Yuhan Cai, † and and Goldberg, D. (2007) 'Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry'. American Chemical Society. doi: 10.1021/AC0617013.

Martínez-Acedo, P. *et al.* (2012) 'A Novel Strategy for Global Analysis of the Dynamic Thiol Redox Proteome', *Molecular & Cellular Proteomics*, 11(9), pp. 800–813. doi: 10.1074/mcp.M111.016469.

Martínez-Bartolomé, S. *et al.* (2008) 'Properties of Average Score Distributions of SEQUEST', *Molecular & Cellular Proteomics*, 7(6), pp. 1135–1145. doi: 10.1074/mcp.M700239-MCP200.

Mattson, M. P. (2010) 'Acetylation Unleashes Protein Demons of Dementia', *Neuron*, 67(6), pp. 900–902. doi: 10.1016/j.neuron.2010.09.010.

Meier, A. and Söding, J. (2015) 'Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling', *PLOS Computational Biology*. Edited by N. Ben-Tal, 11(10), p. e1004343. doi: 10.1371/journal.pcbi.1004343.

Mitchell Wells, J. and McLuckey, S. A. (2005) 'Collision-Induced Dissociation (CID) of Peptides and Proteins', *Methods in Enzymology*. Academic Press, 402, pp. 148–185. doi: 10.1016/S0076-6879(05)02005-7.

Mollapour, M. and Neckers, L. (2012) 'Post-translational modifications of Hsp90 and their contributions to chaperone regulation', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(3), pp. 648–655. doi: 10.1016/j.bbamcr.2011.07.018.

Navarro, P. *et al.* (2014a) 'General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling', *Journal of Proteome Research*. American Chemical Society, 13(3), pp. 1234–1247. doi: 10.1021/pr4006958.

Navarro, P. *et al.* (2014b) 'General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling', *Journal of Proteome Research*. American Chemical Society, 13(3), pp. 1234–1247. doi: 10.1021/pr4006958.

Navarro, P. and Vázquez, J. (2009) 'A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases', *Journal of Proteome Research*, 8(4), pp. 1792–1796. doi: 10.1021/pr800362h.

Nesvizhskii, A. I. (2010) 'A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics', *Journal of Proteomics*, 73(11), pp. 2092–2123. doi: 10.1016/j.jprot.2010.08.009.

Nicholls, S. J. and Hazen, S. L. (2009) 'Myeloperoxidase, modified lipoproteins, and atherogenesis.', *Journal of lipid research*. American Society for Biochemistry and Molecular Biology, 50 Suppl(Suppl), pp. S346-51. doi: 10.1194/jlr.R800086-JLR200.

Oliveira, A. P. and Sauer, U. (2012) 'The importance of post-translational modifications in regulating Saccharomyces cerevisiae metabolism', *FEMS Yeast Research*. Wiley/Blackwell (10.1111), 12(2), pp. 104–117. doi: 10.1111/j.1567-1364.2011.00765.x.

Olsen, J. V and Mann, M. (2013) 'Status of large-scale analysis of post-translational modifications by mass spectrometry.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 12(12), pp. 3444–52. doi: 10.1074/mcp.O113.034181.

Ong, S.-E. and Mann, M. (2007) 'A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)', *Nature Protocols*, 1(6), pp. 2650–2660. doi: 10.1038/nprot.2006.427.

Parker, B. L. *et al.* (2013) 'Site-Specific Glycan-Peptide Analysis for Determination of *N* -Glycoproteome Heterogeneity', *Journal of Proteome Research*, 12(12), pp. 5791–5800. doi: 10.1021/pr400783j.

van Pée, K.-H. and Patallo, E. P. (2006) 'Flavin-dependent halogenases involved in secondary metabolism in bacteria', *Applied Microbiology and Biotechnology*, 70(6), pp. 631–641. doi:

10.1007/s00253-005-0232-2.

Perkins, D. N. *et al.* (1999) 'Probability-based protein identification by searching sequence databases using mass spectrometry data', *Electrophoresis*, 20(18), pp. 3551–3567. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Pirooznia, S. K. and Elefant, F. (2013) 'Targeting specific HATs for neurodegenerative disease treatment: translating basic biology to therapeutic possibilities.', *Frontiers in cellular neuroscience*. Frontiers Media SA, 7, p. 30. doi: 10.3389/fncel.2013.00030.

Potgieter, H. C. *et al.* (1997) 'Spontaneous Oxidation of Methionine: Effect on the Quantification of Plasma Methionine Levels', *Analytical Biochemistry*. Academic Press, 248(1), pp. 86–93. doi: 10.1006/ABIO.1997.2075.

Prabakaran, S. *et al.* (2012) 'Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding.', *Wiley interdisciplinary reviews. Systems biology and medicine*. NIH Public Access, 4(6), pp. 565–83. doi: 10.1002/wsbm.1185.

Rigbolt, K. T. G. *et al.* (2011) 'System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation', *Science Signaling*, 4(164), pp. rs3-rs3. doi: 10.1126/scisignal.2001570.

Roepstorff, P. and Fohlman, J. (1984) 'Letter to the editors', *Biological Mass Spectrometry*, 11(11), pp. 601–601. doi: 10.1002/bms.1200111109.

Ross, P. L. *et al.* (2004) 'Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents', *Molecular & Cellular Proteomics*, 3(12), pp. 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

Roth, G. A. *et al.* (2015) 'Global and Regional Patterns in Cardiovascular Mortality From 1990 to 2013', *Circulation*, 132(17), pp. 1667–1678. doi: 10.1161/CIRCULATIONAHA.114.008720.

Sadayappan, S. *et al.* (2009) 'Cardiac Myosin Binding Protein-C Phosphorylation in a -Myosin Heavy Chain Background', *Circulation*, 119(9), pp. 1253–1262. doi: 10.1161/CIRCULATIONAHA.108.798983.

Santos, A. L. and Lindner, A. B. (2017) 'Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease.', *Oxidative medicine and cellular longevity*. Hindawi Limited, 2017, p. 5716409. doi: 10.1155/2017/5716409.

Savitski, M. M., Nielsen, M. L. and Zubarev, R. A. (2006) 'ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 5(5), pp. 935–48. doi: 10.1074/mcp.T500034-MCP200.

Sharpley, M. S. *et al.* (2012) 'Heteroplasmy of Mouse mtDNA Is Genetically Unstable and Results in Altered Behavior and Cognition', *Cell*, 151(2), pp. 333–343. doi:

10.1016/j.cell.2012.09.004.

Shortreed, M. R. *et al.* (2015) 'Global Identification of Protein Post-translational Modifications in a Single-Pass Database Search', *Journal of Proteome Research*. American Chemical Society, 14(11), pp. 4714–4720. doi: 10.1021/acs.jproteome.5b00599.

Skinner, O. S. and Kelleher, N. L. (2015) 'Illuminating the dark matter of shotgun proteomics', *Nature Biotechnology*. Nature Publishing Group, 33(7), pp. 717–718. doi: 10.1038/nbt.3287.

Stenflo, J. *et al.* (1974) 'Vitamin K dependent modifications of glutamic acid residues in prothrombin.', *Proceedings of the National Academy of Sciences of the United States of America*, 71(7), pp. 2730–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4528109 (Accessed: 7 August 2018).

Tachibana, M. *et al.* (2009) 'Mitochondrial gene replacement in primate offspring and embryonic stem cells', *Nature*. Nature Publishing Group, 461(7262), pp. 367–372. doi: 10.1038/nature08368.

Taus, T. *et al.* (2011) 'Universal and Confident Phosphorylation Site Localization Using phosphoRS', *Journal of Proteome Research*, 10(12), pp. 5354–5362. doi: 10.1021/pr200611n.

Thingholm, T. E. and Larsen, M. R. (2016a) 'Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography', in *Methods in molecular biology (Clifton, N.J.)*, pp. 123–133. doi: 10.1007/978-1-4939-3049-4_8.

Thingholm, T. E. and Larsen, M. R. (2016b) 'The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides', in *Methods in molecular biology (Clifton, N.J.)*, pp. 135–146. doi: 10.1007/978-1-4939-3049-4_9.

Thompson, A. *et al.* (2003) 'Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.', *Analytical chemistry*, 75(8), pp. 1895–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12713048 (Accessed: 21 April 2018).

Wang, Y.-T. *et al.* (2010) 'An Informatics-assisted Label-free Quantitation Strategy that Depicts Phosphoproteomic Profiles in Lung Cancer Cell Invasion', *Journal of Proteome Research*, 9(11), pp. 5582–5597. doi: 10.1021/pr100394u.

Wiśniewski, J. R., Zielinska, D. F. and Mann, M. (2011) 'Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method', *Analytical Biochemistry*, 410(2), pp. 307–309. doi: 10.1016/j.ab.2010.12.004.

Woods, D. and Tilly, J. (2015) 'Autologous Germline Mitochondrial Energy Transfer (AUGMENT) in Human Assisted Reproduction', *Seminars in Reproductive Medicine*, 33(06), pp. 410–421. doi: 10.1055/s-0035-1567826.

Xudong Yao *et al.* (2001) 'Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus'. American Chemical Society. doi: 10.1021/AC001404C.

and, E. S. B. and Fales*, H. M. (2001) 'Overalkylation of a Protein Digest with Iodoacetamide'. American Chemical Society . doi: 10.1021/AC0103423.

Ashkenazy, H. *et al.* (2016) 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.', *Nucleic acids research*. Oxford University Press, 44(W1), pp. W344-50. doi: 10.1093/nar/gkw408.

Bai, B. *et al.* (2017) 'Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry.', *Methods in enzymology*. NIH Public Access, 585, pp. 377–395. doi: 10.1016/bs.mie.2016.10.007.

Ballif, B. A. *et al.* (2004) 'Phosphoproteomic Analysis of the Developing Mouse Brain', *Molecular & Cellular Proteomics*, 3(11), pp. 1093–1101. doi: 10.1074/mcp.M400085-MCP200.

Batth, T. S., Francavilla, C. and Olsen, J. V. (2014) 'Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics', *Journal of Proteome Research*. American Chemical Society, 13(12), pp. 6176–6186. doi: 10.1021/pr500893m.

Bjornson, R. D. *et al.* (2008) 'X!!Tandem, an Improved Method for Running X!Tandem in Parallel on Collections of Commodity Computers', *Journal of Proteome Research*, 7(1), pp. 293–299. doi: 10.1021/pr0701198.

Bonzon-Kulichenko, E. *et al.* (2015) 'Revisiting Peptide Identification by High-Accuracy Mass Spectrometry: Problems Associated with the Use of Narrow Mass Precursor Windows', *Journal of Proteome Research*. American Chemical Society, 14(2), pp. 700–710. doi: 10.1021/pr5007284.

Broncel, M. *et al.* (2010) 'How Post-Translational Modifications Influence Amyloid Formation: A Systematic Study of Phosphorylation and Glycosylation in Model Peptides', *Chemistry - A European Journal*. Wiley-Blackwell, 16(26), pp. 7881–7888. doi: 10.1002/chem.200902452.

Carapito, C. *et al.* (2009) 'Systematic LC-MS Analysis of Labile Post-Translational Modifications in Complex Mixtures', *Journal of Proteome Research*, 8(5), pp. 2608–2614. doi: 10.1021/pr800871n.

Chalkley, R. J. *et al.* (2008) 'In-depth Analysis of Tandem Mass Spectrometry Data from Disparate Instrument Types', *Molecular & Cellular Proteomics*, 7(12), pp. 2386–2398. doi: 10.1074/mcp.M800021-MCP200.

Chammas, R. *et al.* (1999) 'De-N-acetyl-gangliosides in humans: unusual subcellular distribution of a novel tumor antigen.', *Cancer research*, 59(6), pp. 1337–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10096568 (Accessed: 8 May 2018).

Chee Sian Gan *et al.* (2007) 'Technical, Experimental, and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ)'. American Chemical Society. doi: 10.1021/PR060474I.

Chen, Y. *et al.* (2009) 'PTMap--a sequence alignment software for unrestricted, accurate, and

full-spectrum identification of post-translational modification sites.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(3), pp. 761–6. doi: 10.1073/pnas.0811739106.

Chi, A. *et al.* (2007) 'Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(7), pp. 2193–8. doi: 10.1073/pnas.0607084104.

Chick, J. M. *et al.* (2015) 'A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides.', *Nature biotechnology*. NIH Public Access, 33(7), pp. 743–9. doi: 10.1038/nbt.3267.

Choi, H., Fermin, D. and Nesvizhskii, A. I. (2008) 'Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics', *Molecular & Cellular Proteomics*, 7(12), pp. 2373–2385. doi: 10.1074/mcp.M800203-MCP200.

Choi, H. and Nesvizhskii, A. I. (2008) 'False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics', *Journal of Proteome Research*, 7(1), pp. 47–50. doi: 10.1021/pr700747q.

Clough, T. *et al.* (2009) 'Protein Quantification in Label-Free LC-MS Experiments', *Journal of Proteome Research*. American Chemical Society, 8(11), pp. 5275–5284. doi: 10.1021/pr900610q.

Cox, J. *et al.* (2011) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*. Nature Publishing Group, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Dai, J. *et al.* (2007) 'Protein Phosphorylation and Expression Profiling by Yin-Yang Multidimensional Liquid Chromatography (Yin-Yang MDLC) Mass Spectrometry', *Journal of Proteome Research*, 6(1), pp. 250–262. doi: 10.1021/pr0604155.

Daly, D. S. *et al.* (2008) 'Mixed-Effects Statistical Model for Comparative LC−MS Proteomics Studies †', *Journal of Proteome Research*. American Chemical Society, 7(3), pp. 1209–1217. doi: 10.1021/pr070441i.

Delporte, C. *et al.* (2013) 'Low-density lipoprotein modified by myeloperoxidase in inflammatory pathways and clinical studies.', *Mediators of inflammation*. Hindawi, 2013, p. 971579. doi: 10.1155/2013/971579.

Diament, B. J. and Noble, W. S. (2011) 'Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra', *J. Proteome Res*, 10, pp. 3871–3879. doi: 10.1021/pr101196n.

Duan, G. and Walther, D. (2015) 'The Roles of Post-translational Modifications in the Context of Protein Interaction Networks', *PLOS Computational Biology*. Edited by P. Radivojac. Public Library of Science, 11(2), p. e1004049. doi: 10.1371/journal.pcbi.1004049.

Elias, J. E. and Gygi, S. P. (2007) 'Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry', *Nature Methods*, 4(3), pp. 207–214. doi: 10.1038/nmeth1019.

Eng, J. K. *et al.* (2008) 'A Fast SEQUEST Cross Correlation Algorithm', *Journal of Proteome Research*. American Chemical Society, 7(10), pp. 4598–4602. doi: 10.1021/pr800420s.

Eng, J. K. *et al.* (2015) 'A Deeper Look into Comet—Implementation and Features', *Journal of The American Society for Mass Spectrometry*, 26(11), pp. 1865–1874. doi: 10.1007/s13361-015-1179-x.

Eng, J. K., Jahan, T. A. and Hoopmann, M. R. (2013a) 'Comet: An open-source MS/MS sequence database search tool', *PROTEOMICS*, 13(1), pp. 22–24. doi: 10.1002/pmic.201200439.

Eng, J. K., Jahan, T. A. and Hoopmann, M. R. (2013b) 'Comet: An open-source MS/MS sequence database search tool', *PROTEOMICS*. Wiley-Blackwell, 13(1), pp. 22–24. doi: 10.1002/pmic.201200439.

Eng, J. K., McCormack, A. L. and Yates, J. R. (1994) 'An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database', *Journal of the American Society for Mass Spectrometry*, 5(11), pp. 976–989. doi: 10.1016/1044-0305(94)80016-2.

Engholm-Keller, K. *et al.* (2012) 'TiSH — a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC', *Journal of Proteomics*, 75(18), pp. 5749–5761. doi: 10.1016/j.jprot.2012.08.007.

Ferries, S. *et al.* (2017) 'Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer', *Journal of Proteome Research*. American Chemical Society, 16(9), pp. 3448–3459. doi: 10.1021/acs.jproteome.7b00337.

Fu, Y. (2012) 'Bayesian false discovery rates for post-translational modification proteomics *', *Statistics and Its Interface*, 5, pp. 47–59. Available at: http://www.amt.ac.cn/member/fuyan/pubs/SII-5-1-a05-Fu.pdf (Accessed: 19 April 2018).

Fu, Y. and Qian, X. (2014) 'Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 13(5), pp. 1359–68. doi: 10.1074/mcp.O113.030189.

Gajjala, P. R. *et al.* (2015) 'Emerging role of post-translational modifications in chronic kidney disease and cardiovascular disease', *Nephrology Dialysis Transplantation*, 30(11), pp. 1814–1824. doi: 10.1093/ndt/gfv048.

García-Marqués, F. *et al.* (2016) 'A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 15(5), pp. 1740–60. doi: 10.1074/mcp.M115.055905.

Griss, J. *et al.* (2016) 'Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets', *Nature Methods*. Nature Publishing Group, 13(8), pp. 651–656. doi: 10.1038/nmeth.3902.

Gupta, M. K. and Robbins, J. (2014) 'Post-translational control of cardiac hemodynamics through myosin binding protein C', *Pflügers Archiv - European Journal of Physiology*, 466(2), pp. 231–236. doi: 10.1007/s00424-013-1377-y.

Hamalainen, R. H. *et al.* (2013) 'Tissue- and cell-type-specific manifestations of heteroplasmic mtDNA 3243A>G mutation in human induced pluripotent stem cell-derived disease model', *Proceedings of the National Academy of Sciences*, 110(38), pp. E3622–E3630. doi: 10.1073/pnas.1311660110.

Hämäläinen, R. H. *et al.* (2015) 'mtDNA Mutagenesis Disrupts Pluripotent Stem Cell Function by Altering Redox Signaling.', *Cell reports*. Elsevier, 11(10), pp. 1614–24. doi: 10.1016/j.celrep.2015.05.009.

Hegab, Z. *et al.* (2012) 'Role of advanced glycation end products in cardiovascular disease.', *World journal of cardiology*. Baishideng Publishing Group Inc, 4(4), pp. 90–102. doi: 10.4330/wjc.v4.i4.90.

Hogrebe, A. *et al.* (2018) 'Benchmarking common quantification strategies for large-scale phosphoproteomics', *Nature Communications*. Nature Publishing Group, 9(1), p. 1045. doi: 10.1038/s41467-018-03309-6.

Hsieh, E. J. *et al.* (2010) 'Comparison of database search strategies for high precursor mass accuracy MS/MS data.', *Journal of proteome research*. NIH Public Access, 9(2), pp. 1138–43. doi: 10.1021/pr900816a.

Hu, C.-W. *et al.* (2015) 'Temporal Phosphoproteome Dynamics Induced by an ATP Synthase Inhibitor Citreoviridin', *Molecular & Cellular Proteomics*, 14(12), pp. 3284–3298. doi: 10.1074/mcp.M115.051383.

Huang, K. *et al.* (2017) 'Proteogenomic integration reveals therapeutic targets in breast cancer xenografts', *Nature Communications*, 8, p. 14864. doi: 10.1038/ncomms14864.

Hultin-Rosenberg, L. *et al.* (2013) 'Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 12(7), pp. 2021–31. doi: 10.1074/mcp.M112.021592.

Huttlin, E. L. *et al.* (2010) 'A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression', *Cell*, 143(7), pp. 1174–1189. doi: 10.1016/j.cell.2010.12.001.

Jenuth, J. P. *et al.* (1996) 'Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA', *Nature Genetics*, 14(2), pp. 146–151. doi: 10.1038/ng1096-146.

Jeong, K., Kim, S. and Bandeira, N. (2012) 'False discovery rates in spectral identification.', *BMC bioinformatics*. BioMed Central, 13 Suppl 16(Suppl 16), p. S2. doi: 10.1186/1471-2105-

13-S16-S2.

Jin, W.-H. *et al.* (2004) 'Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry', *Rapid Communications in Mass Spectrometry*, 18(18), pp. 2169–2176. doi: 10.1002/rcm.1604.

Junmin Peng, † *et al.* (2002) 'Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome'. American Chemical Society. doi: 10.1021/PR025556V.

Karpievitch, Y. *et al.* (2009) 'A statistical framework for protein quantitation in bottom-up MS-based proteomics', *Bioinformatics*, 25(16), pp. 2028–2034. doi: 10.1093/bioinformatics/btp362.

Keller, A. *et al.* (2002) 'Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.', *Analytical chemistry*, 74(20), pp. 5383–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12403597 (Accessed: 26 April 2018).

Khadjavi, A. *et al.* (2011) 'Evidence of Abnormal Tyrosine Phosphorylated Proteins in the Urine of Patients With Bladder Cancer: The Road Toward a New Diagnostic Tool?', *The Journal of Urology*, 185(5), pp. 1922–1929. doi: 10.1016/j.juro.2010.12.029.

Kim, J.-S. *et al.* (2013) 'In-Source Fragmentation and the Sources of Partially Tryptic Peptides in Shotgun Proteomics', *Journal of Proteome Research*, 12(2), pp. 910–916. doi: 10.1021/pr300955f.

Kim, S. and Pevzner, P. A. (2014) 'MS-GF+ makes progress towards a universal database search tool for proteomics', *Nature Communications*. Nature Publishing Group, 5, p. 5277. doi: 10.1038/ncomms6277.

Klammer, A. A., Park, C. Y. and Noble, W. S. (2009) 'Statistical calibration of the SEQUEST Xcorr function', *Journal of Proteome Research*, 8(4), pp. 2106–2113. doi: 10.1021/pr8011107.

Kong, A. T. *et al.* (2017) 'MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics', *Nature Methods*. Nature Publishing Group, 14(5), pp. 513–520. doi: 10.1038/nmeth.4256.

Lee, Y. *et al.* (2014) 'Compromised Catalysis and Potential Folding Defects in *in Vitro* Studies of Missense Mutants Associated with Hereditary Phosphoglucomutase 1 Deficiency', *Journal of Biological Chemistry*, 289(46), pp. 32010–32019. doi: 10.1074/jbc.M114.597914.

Leyfer, D. and Weng, Z. (2005) 'Genome-wide decoding of hierarchical modular structure of transcriptional regulation by cis-element and expression clustering', *Bioinformatics*, 21(Suppl 2), pp. ii197-ii203. doi: 10.1093/bioinformatics/bti1131.

Li, Q. *et al.* (2017) 'Global Post-Translational Modification Discovery.', *Journal of proteome research*. American Chemical Society, 16(4), pp. 1383–1390. doi: 10.1021/acs.jproteome.6b00034.

Lina P. Aristoteli, †, Mark P. Molloy, †, ‡ and and Mark S. Baker*, †, ‡ (2006) 'Evaluation of Endogenous Plasma Peptide Extraction Methods for Mass Spectrometric Biomarker

Discovery'. American Chemical Society. doi: 10.1021/PR0602996.

Link, A. J. *et al.* (1999) 'Direct analysis of protein complexes using mass spectrometry', *Nature Biotechnology*. Nature Publishing Group, 17(7), pp. 676–682. doi: 10.1038/10890.

Liu, L. and Pilch, P. F. (2016) 'PTRF/Cavin-1 promotes efficient ribosomal RNA transcription in response to metabolic challenges', *eLife*. eLife Sciences Publications Limited, 5, p. e17508. doi: 10.7554/eLife.17508.

Lohrig, K., Sickmann, A. and Lewandrowski, U. (2011) 'Strong Cation Exchange Chromatography for Analysis of Sialylated Glycopeptides', in *Methods in molecular biology (Clifton, N.J.)*, pp. 299–308. doi: 10.1007/978-1-61779-148-2_20.

López-Carballo, G. *et al.* (2002) 'Activation of the Phosphatidylinositol 3-Kinase/Akt Signaling Pathway by Retinoic Acid Is Required for Neural Differentiation of SH-SY5Y Human Neuroblastoma Cells', *Journal of Biological Chemistry*, 277(28), pp. 25297–25304. doi: 10.1074/jbc.M201869200.

Lu, B. *et al.* (2007) 'Automatic Validation of Phosphopeptide Identifications from Tandem Mass Spectra', *Analytical Chemistry*, 79(4), pp. 1301–1310. doi: 10.1021/ac061334v.

Lundby, A. *et al.* (2013) 'In Vivo Phosphoproteomics Analysis Reveals the Cardiac Targets of -Adrenergic Receptor Signaling', *Science Signaling*, 6(278), pp. rs11-rs11. doi: 10.1126/scisignal.2003506.

Ma, C. W. M. and Lam, H. (2014) 'Hunting for Unexpected Post-Translational Modifications by Spectral Library Searching with Tier-Wise Scoring', *Journal of Proteome Research*. American Chemical Society, 13(5), pp. 2262–2271. doi: 10.1021/pr401006g.

Mann, M. and Jensen, O. N. (2003) 'Proteomic analysis of post-translational modifications', *Nature Biotechnology*. Nature Publishing Group, 21(3), pp. 255–261. doi: 10.1038/nbt0303-255.

Marshall Bern, *, Yuhan Cai, † and and Goldberg, D. (2007) 'Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry'. American Chemical Society. doi: 10.1021/AC0617013.

Martínez-Acedo, P. *et al.* (2012) 'A Novel Strategy for Global Analysis of the Dynamic Thiol Redox Proteome', *Molecular & Cellular Proteomics*, 11(9), pp. 800–813. doi: 10.1074/mcp.M111.016469.

Martínez-Bartolomé, S. *et al.* (2008) 'Properties of Average Score Distributions of SEQUEST', *Molecular & Cellular Proteomics*, 7(6), pp. 1135–1145. doi: 10.1074/mcp.M700239-MCP200.

Mattson, M. P. (2010) 'Acetylation Unleashes Protein Demons of Dementia', *Neuron*, 67(6), pp. 900–902. doi: 10.1016/j.neuron.2010.09.010.

Meier, A. and Söding, J. (2015) 'Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling', *PLOS Computational Biology*. Edited by N. Ben-Tal, 11(10), p. e1004343. doi: 10.1371/journal.pcbi.1004343.

Mitchell Wells, J. and McLuckey, S. A. (2005) 'Collision-Induced Dissociation (CID) of Peptides and Proteins', *Methods in Enzymology*. Academic Press, 402, pp. 148–185. doi: 10.1016/S0076-6879(05)02005-7.

Mollapour, M. and Neckers, L. (2012) 'Post-translational modifications of Hsp90 and their contributions to chaperone regulation', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(3), pp. 648–655. doi: 10.1016/j.bbamcr.2011.07.018.

Navarro, P. *et al.* (2014a) 'General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling', *Journal of Proteome Research*. American Chemical Society, 13(3), pp. 1234–1247. doi: 10.1021/pr4006958.

Navarro, P. *et al.* (2014b) 'General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling', *Journal of Proteome Research*. American Chemical Society, 13(3), pp. 1234–1247. doi: 10.1021/pr4006958.

Navarro, P. and Vázquez, J. (2009) 'A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases', *Journal of Proteome Research*, 8(4), pp. 1792–1796. doi: 10.1021/pr800362h.

Nesvizhskii, A. I. (2010) 'A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics', *Journal of Proteomics*, 73(11), pp. 2092–2123. doi: 10.1016/j.jprot.2010.08.009.

Nicholls, S. J. and Hazen, S. L. (2009) 'Myeloperoxidase, modified lipoproteins, and atherogenesis.', *Journal of lipid research*. American Society for Biochemistry and Molecular Biology, 50 Suppl(Suppl), pp. S346-51. doi: 10.1194/jlr.R800086-JLR200.

Oliveira, A. P. and Sauer, U. (2012) 'The importance of post-translational modifications in regulating Saccharomyces cerevisiae metabolism', *FEMS Yeast Research*. Wiley/Blackwell (10.1111), 12(2), pp. 104–117. doi: 10.1111/j.1567-1364.2011.00765.x.

Olsen, J. V and Mann, M. (2013) 'Status of large-scale analysis of post-translational modifications by mass spectrometry.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 12(12), pp. 3444–52. doi: 10.1074/mcp.O113.034181.

Ong, S.-E. and Mann, M. (2007) 'A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)', *Nature Protocols*, 1(6), pp. 2650–2660. doi: 10.1038/nprot.2006.427.

Parker, B. L. *et al.* (2013) 'Site-Specific Glycan-Peptide Analysis for Determination of *N* -Glycoproteome Heterogeneity', *Journal of Proteome Research*, 12(12), pp. 5791–5800. doi: 10.1021/pr400783j.

van Pée, K.-H. and Patallo, E. P. (2006) 'Flavin-dependent halogenases involved in secondary metabolism in bacteria', *Applied Microbiology and Biotechnology*, 70(6), pp. 631–641. doi: 10.1007/s00253-005-0232-2.

Perkins, D. N. *et al.* (1999) 'Probability-based protein identification by searching sequence databases using mass spectrometry data', *Electrophoresis*, 20(18), pp. 3551–3567. doi:

10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Pirooznia, S. K. and Elefant, F. (2013) 'Targeting specific HATs for neurodegenerative disease treatment: translating basic biology to therapeutic possibilities.', *Frontiers in cellular neuroscience*. Frontiers Media SA, 7, p. 30. doi: 10.3389/fncel.2013.00030.

Potgieter, H. C. *et al.* (1997) 'Spontaneous Oxidation of Methionine: Effect on the Quantification of Plasma Methionine Levels', *Analytical Biochemistry*. Academic Press, 248(1), pp. 86–93. doi: 10.1006/ABIO.1997.2075.

Prabakaran, S. *et al.* (2012) 'Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding.', *Wiley interdisciplinary reviews. Systems biology and medicine*. NIH Public Access, 4(6), pp. 565–83. doi: 10.1002/wsbm.1185.

Rigbolt, K. T. G. *et al.* (2011) 'System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation', *Science Signaling*, 4(164), pp. rs3-rs3. doi: 10.1126/scisignal.2001570.

Roepstorff, P. and Fohlman, J. (1984) 'Letter to the editors', *Biological Mass Spectrometry*, 11(11), pp. 601–601. doi: 10.1002/bms.1200111109.

Ross, P. L. *et al.* (2004) 'Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents', *Molecular & Cellular Proteomics*, 3(12), pp. 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

Roth, G. A. *et al.* (2015) 'Global and Regional Patterns in Cardiovascular Mortality From 1990 to 2013', *Circulation*, 132(17), pp. 1667–1678. doi: 10.1161/CIRCULATIONAHA.114.008720.

Sadayappan, S. *et al.* (2009) 'Cardiac Myosin Binding Protein-C Phosphorylation in a -Myosin Heavy Chain Background', *Circulation*, 119(9), pp. 1253–1262. doi: 10.1161/CIRCULATIONAHA.108.798983.

Santos, A. L. and Lindner, A. B. (2017) 'Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease.', *Oxidative medicine and cellular longevity*. Hindawi Limited, 2017, p. 5716409. doi: 10.1155/2017/5716409.

Savitski, M. M., Nielsen, M. L. and Zubarev, R. A. (2006) 'ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 5(5), pp. 935–48. doi: 10.1074/mcp.T500034-MCP200.

Shortreed, M. R. *et al.* (2015) 'Global Identification of Protein Post-translational Modifications in a Single-Pass Database Search', *Journal of Proteome Research*. American Chemical Society, 14(11), pp. 4714–4720. doi: 10.1021/acs.jproteome.5b00599.

Silva, M. L. S. (2015) 'Cancer serum biomarkers based on aberrant post-translational modifications of glycoproteins: Clinical value and discovery strategies', *Biochimica et*

Biophysica Acta (BBA) - Reviews on Cancer. Elsevier, 1856(2), pp. 165–177. doi: 10.1016/J.BBCAN.2015.07.002.

Skinner, O. S. and Kelleher, N. L. (2015) 'Illuminating the dark matter of shotgun proteomics', *Nature Biotechnology*. Nature Publishing Group, 33(7), pp. 717–718. doi: 10.1038/nbt.3287.

Thingholm, T. E. and Larsen, M. R. (2016a) 'Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography', in *Methods in molecular biology (Clifton, N.J.)*, pp. 123–133. doi: 10.1007/978-1-4939-3049-4_8.

Thingholm, T. E. and Larsen, M. R. (2016b) 'The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides', in *Methods in molecular biology (Clifton, N.J.)*, pp. 135–146. doi: 10.1007/978-1-4939-3049-4_9.

Thompson, A. *et al.* (2003) 'Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.', *Analytical chemistry*, 75(8), pp. 1895–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12713048 (Accessed: 21 April 2018).

Wang, Y.-T. *et al.* (2010) 'An Informatics-assisted Label-free Quantitation Strategy that Depicts Phosphoproteomic Profiles in Lung Cancer Cell Invasion', *Journal of Proteome Research*, 9(11), pp. 5582–5597. doi: 10.1021/pr100394u.

Wiśniewski, J. R., Zielinska, D. F. and Mann, M. (2011) 'Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method', *Analytical Biochemistry*, 410(2), pp. 307–309. doi: 10.1016/j.ab.2010.12.004.

Xudong Yao *et al.* (2001) 'Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus'. American Chemical Society. doi: 10.1021/AC001404C.

Appendices

1. Appendix 1: Supplementary tables

Due to the multiple big files, the links are provided. In addition, a copy of Tables in excel sheet forms, has been provided in CD

1.1. Table 1: Benchmarking of Comet-PTM using synthetic Phospho-peptide

Comet-PTM correctly localized the position of the modified residue in 81% (122 out of 151) of the cases. In 94% of the cases (144 out of 151), the site was localized in the correct or in an adjacent position.

Link to download file:

https://www.cell.com/cms/attachment/2119363229/2092777175/mmc2.xlsx

1.2. Table 2: List of functional categories significantly altered by heteroplasmy (FDR < 5%) as a consequence of a coordinated protein response

Increases in the heteroplasmic tissues, in relation to the average protein abundance of all samples are coloured in red, whereas decreases are in blue, according to the graded scale.

Link to download file:

https://www.cell.com/cms/attachment/2119363229/2092777176/mmc3.xlsx

1.3. Table 3: List of modified peptides significantly altered by heteroplasmy (p < 0.05)

Increases in peptide abundance in the heteroplasmic tissue compared to the average of peptide abundance of all samples are coloured in red, whereas decreases are in blue, according to the graded scale. P values are calculated using Student's t test.

Link to download file:

https://www.cell.com/cms/attachment/2119363229/2092777177/mmc4.xlsx

1.4. Table 4: Conservation and structural analysis of heteroplasmy-modified peptides of OXPHOS complexes I, III, IV, and V

The residues harbouring the modification are underlined. Residues predicted to be functional (conservation score 8-9 and exposed) are highlighted in red; residues predicted to have structural implications (conservation score 9 and buried), in blue; highly conserved residues (conservation score > 5), in green. Most PTMs are on highly conserved residues or next to highly conserved residues. Conservation scores were rated from 9 (highest conservation) to 1 (highest variability) by the ConSurf server

Link to download file:

https://www.cell.com/cms/attachment/2119363229/2092777178/mmc5.xlsx

2. Appendix 2: Help for the programs developed

2.1. Comet-PTM

Comet-PTM is a modified version of the Comet project for the thorough analysis of post-translational modifications, developed at the Cardiovascular Proteomics laboratory at CNIC (Spain). Link to download: https://github.com/CNIC-Proteomics/Comet-PTM

Here i will describe the details that are different in Comet-PTM, especially the Comet-PTM specific parameters and its output. For the remaining parameters and output common with the original version of Comet, you have complete documentation at http://comet-ms.sourceforge.net/

Input information:

```
#
# CNIC / comet-iq / comet-PTM specific
#
# do deltaX search for: (-delta_tolerance_outer < deltaMass < -delta_tolerance_inner) OR
#                       (+delta_tolerance_inner < deltaMass < +delta_tolerance_outer)
#
use_delta_xcorr          = 1     # 0=no (default), 1=yes
delta_outer_tolerance    = 500   # ignored if use_delta_xcorr 0, default 320
delta_inner_tolerance    = 0     # ignored if use_delta_xcorr 0, default 0.8
use_delta_back_jumps     = 0     # 0=no (default), 1=yes
use_delta_forward_jumps  = 0     # 0=no (default), 1=yes
dont_calc_pseudo_non_mod = 1     # 0=yes, calculate pseudo nonmod (default), 1=avoid calculating pseudo non_mod
```

- *use_delta_xcorr*, checks whether the DeltaXCorr (the mass difference between the theoretical mass of the peptide identified and the experimental mass of potentially modified peptides) will be used or not (this is needed to use Comet-PTM).
- *delta_outer_tolerance*, the tolerance in Da for the maximum Delta XCorr that should be considered to calculate the mass of the modified peptide.
- *delta_inner_tolerance*, the minimum value of the Delta Xcorr, under which a mass difference is not taken into consideration (for being considered a non-modified peptide).
- *use_delta_back_jumps*, to consider isotopologues *after* the peak analysed.
- *use_delta_forward_jumps*, to consider isotopologues *before* the peak analysed.
- *dont_calc_pseudo_non_mod*, by default the nonmodified peptide is included among search results; by activating this parameter, those peptides will not be presented.

Output information:

Depending on the options in the params file, the output information can be in a number of formats. For handling information we recommend using the tab-separated tables generated by setting output_txtfile = 1

The generated file contains a number of columns. The following columns are the same as the columns used in regular comet: *scan, num, charge, exp_neutral_mass, calc_neutral_mass, e-value, xcorr, delta_cn, sp_score, ions_matched, ions_total, plain_peptide, retention_time, modifications* (in regular comet this column appears at the end), *prev_aa, next_aa, protein, protein_count*. The column *modified_peptide* does not appear as such in Comet-PTM.

The columns described below are Comet-PTM specific:

- *exp_mz*, the experimental *m/z* calculated as *exp_neutral_mass - proton_mass * z*.
- *q_score*, dot product of intensities of matched ions.
- *peptide*, same as *plain_peptide*, but including previous and posterior amino acids.
- *delta_mods*, the mass difference between the theoretical (non-modified) peptide and the modified peptide calculated and assigned to the residue, calculated as *exp_neutral_mass - calc_neutral_mass*.
- *b_series_delta_mods*, the fragments of the b-series that have been observed to have a displacement of mass *delta_mods*. Every fragment that has been observed with such mass shift is followed by the "<" character.
- *y_series_delta_mods*, sama as previous, but for the y-series fragments. Every fragment that has been observed with such mass shift is preceded by the ">" character.
- *delta_jumps*, can be '-2', '-1', '0', '+1', '+2', and shows the backward/forward correction to select the monoisotopic peak from the isotopic envelope (deprecated, for regular options it always shows '0').
- *delta_peptide*, shows the sequence presented in *plain_peptide*, but including the mass shift observed in *delta_mods* in brackets.

2.2. SHIFTS

SHIFTS is a program made in the Jesus Vazquez Cardiovascular Proteomics Lab at Centro Nacional de Investigaciones Cardiovasculares (CNIC), for high throughput PTM (Post translation modifications) processing SHIFTS identifies peaks in the Delta Mass distribution, assigns PSM to peaks and calculates FDR for peptide identification. SHIFTS is developed in python programming language as command line tool and its works with cometPTM produced ".txt" files. Link to download: https://github.com/CNIC-Proteomics/SHIFTS

Input information:

There are eight input parameters for SHIFTS.

```
-P,        --Path2master

                --Input will be a text file, containing two columns: first, would be
                master path and second as sub path. This parameter will help to analyse
                multiple experiment of a project together, i.e. Project name = PTM,
                experiments in project PTM: heart, liver. Therefore, path/PTM will be
                master Path and heart and liver will be sub path.

-B,        --BinSize

                --Bin size for slope modelling, the most standard bin to use is 0.001
                which has been tested for several experiments, however you are free to
                choose depending on your machine resolution. Lower the resolution
                bigger the bin.

-X,        --Xcor

                --Corrected Xcorr threshold for choosing the best non-modified
                peptides, based on provided Xcorr, mass calibration will be performed.
                In most cases, 0.20 or 0.25 would be good enough.

-F,        --Fastafile

                --path to Protein database which was used for searches (Concatenated).

-A,        --ApexFilter

                --Apex threshold, a simple integer input (example 10 0r 15 0r 8) for
                ignoring the PTM peaks which have less PSMs than specified number. It
                will help to avoid the background noise. However in cases of small
                experiments using 0 would be useful, as some good PTMs will be very
                low frequent.

-f,        --FDRthreshold

                --Global FDR filtration threshold. Example: 0.01 0r 0.05.
```

-O, --OutPutname

--output folder name, specified name will be created in every experiment folder with all the results file.

-i, --isotopeCorrection

--it´s True/false condition for isotopic correction. 0 for false and 1 for true. A simple isotopic correction to minimize miss assignations of the correct monoisotopic peak of the precursor. When two PSM having the same sequence are encountered having a Mass difference within 1 ppm of the mass difference expected for either one or two 13C or one 34S, the ?Mass of the heaviest precursor is substituted by that of the lightest one.

-l, --localFDR

--Local FDR filtration threshold. Example: 0.01 0r 0.05.

-p, --localFDR

--Peak FDR filtration threshold. Example: 0.01 0r 0.05.

Output information:

1)"experimentName_TargetData_Calibration.txt"

--This is the first file to be created by SHIFTS, and will contain all the fractions from an experiment with the calibration error per fraction with recalibrated masses. From here to the results, this file will be used by SHIFTS internally.

2)"GlobalFDR.txt"

--This file will contain the global FDR Xcorr threshold for specified FDR in input parameter "-f".

3)"sigmaCalculations.txt"

--This file will contain the fitting value for sigma.

4)"heart_Median_log.txt"

--This file contain all the calibration error value, and number of PSM/Fraction from which error has been calculated.

5)"target_Peak_identification_histogram.txt"

--This file contains a histogram and Gaussian derivatives for all the delta masses, which are used for peak apex picking.

6)"Decoy_Peak_identification_histogram.txt"

--This file contains a histogram from decoys.

7)"SlopeFDRfile.txt"

--This file contains all the data, like the calibration file, but with Local FDR calculated with every 1 Da.

8)"Peak_and_Slope_FDRfile.txt"

--This file contains all the data, like the SlopeFDR file, but with Peak FDR, calculated with every peak apex.

9)"NotassignedSequences.txt"

--This file contains some duplicated (for computing convenience) information, which will be found in final file.

10)"AllWithSequence-massTag.txt"

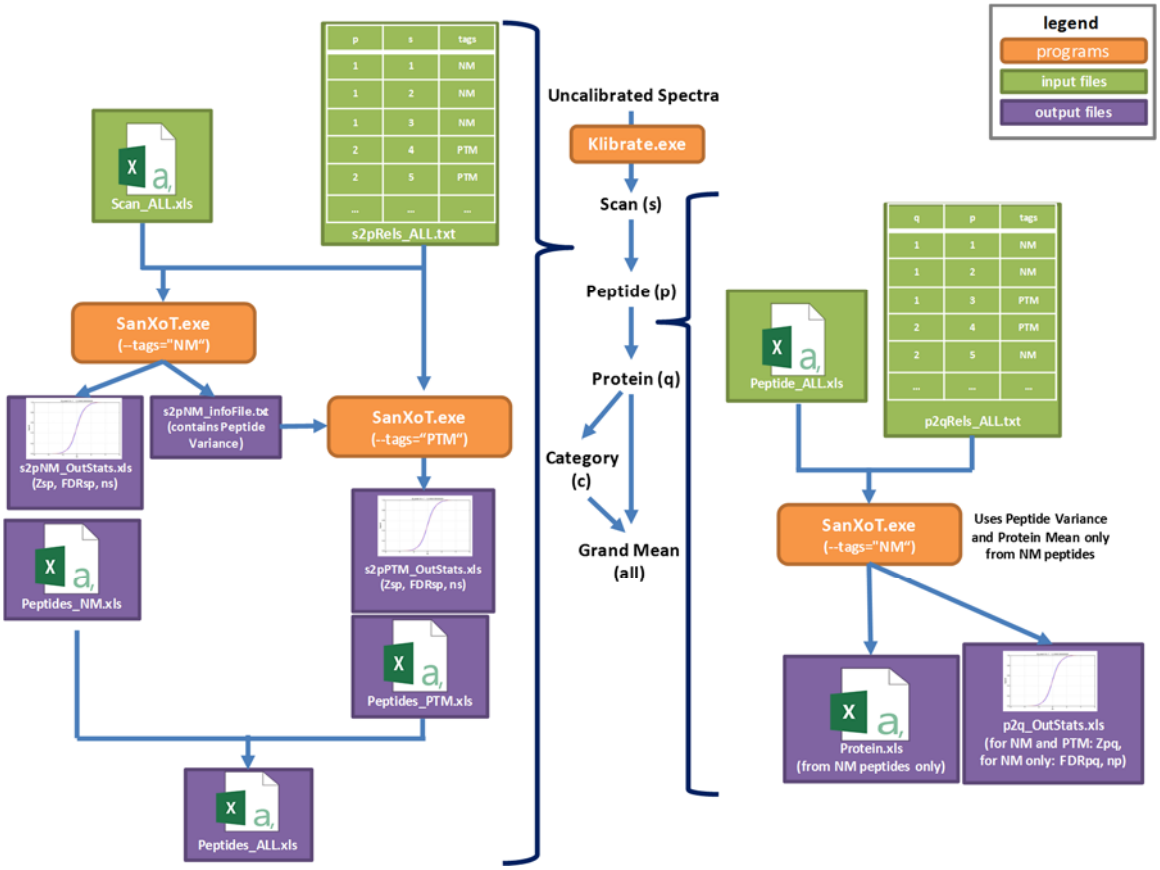
--This file contains all the data with FDR filtration and PSM to peak apex tagging. (PSM to Peak assignment). This will be a final File in case of where -i from parameter defined as 0.

11)"IsotopCorrection_TargetData_withSequence-massTag.txt"

--Will be the final file in case of where -i is defined as 1. Contains all the PSMs after FDR filter and with isotopic corrections.

2.3.PTM-Quant-Stats

The GIA workflows are constructed using three executables: klibrate.exe, sanxot.exe and sanxotsieve.exe. Workflows are assembled in .bat files that are executed automatically under Windows command prompt (the same .bat file was used to perform the results provided in this thesis). All the executables and detailed help is provided online with example dataset. Link to download the pipeline and data: https://github.com/CNIC-Proteomics/PTM-Quant-Stats



Supplementary Figure 1: representation quantification workflow, in which the automated pipeline had been executed

Detail explanation of .Bat file, used in this work.

```
REMImportant: before performing the integration, change "full path" by the path containing
data and relation tables
```

```
REM*********************************************
REM********** PART 1: integrations ************
REM******** scan > peptide > protein ***********
REM*********************************************
REM********** CALIBRATION OF WEIGHTS (as in the WSPP model)
```

```
REMCalculation of calibration constant k of the set of scans belonging to the NM-peptides
REMk transforms un-calibrated weights into true weights (inverses of variances)
REM-a sets the indicated prefix into the output file
```

REM-o names the output file as "NM_scan_Calibrated.txt"

klibrate.exe -d"full_path/scan_XsVs_NM.txt" -r"full_path/s2pNM_Rels.txt" -k1 -v0 -aNM_scan -o"full_path/NM_scan_Calibrated.txt"

REMCalibration of weights of the set of scans belonging to the PTM-peptides

REM-f avoids k and variance recalculation (uses variance seed as final variance)

REMthe variance and calibration constant k from the set of scans corresponding to the non-modified (NM) peptides are used to calibrate the scans from the PTM-containing peptides (PTM)

klibrate.exe -d"full_path/scan_XsVs_PTM.txt" -full_path/s2pPTM_Rels.txt" –f -K"full_path/NM_scan_infoFile.txt" -V"full_path/NM_scan_infoFile.txt" -aPTM_scan -o"full_path/PTM_scan_Calibrated.txt"

REMprepares a joined scan data file from NM and PTM peptides

Copy "full_path/NM_scan_Calibrated.txt"+"full_path/PTM_scan_Calibrated.txt" "full_path/ALL_scan.txt"

REMprepares a joined scan to peptide relations file from NM and PTM peptides

copy "full_path/s2pNM_Rels.txt"+"full_path/s2pPTM_Rels.txt" "full_path/s2pALL_Rels.txt"

REM******** SCAN TO PEPTIDE INTEGRATION

REM--tags"NonMod" selects the group of scans from the NM peptides as the null-hypothesis to calculate the scan variance.

REMThe tag is included in the third column of the scan-to-peptide relations file.

REMA file named "s2pNM_OutStats.xls" is created. It contains Zsp for scans from NM and PTM-peptides, and FDRsp for scans only from NM peptides.

REMAlso a file named "NM_peptide.xls" is created. It contains Xp and Vp values from NM-peptides to perform the following peptide-to-protein integration.

sanxot.exe -d"full_path/ALL_scan.txt" -r"full_path/s2pALL_Rels.txt" -a"s2pNM" -o"NM_peptide.xls" --tags"NonMod"

REM-f avoids variance recalculation (uses variance seed as final variance)

REM-V"s2pNM_infoFile.txt" forces the previously obtained null-hypothesis scan variance as seed to integrate the scans from the PTM-peptides.

REM--tags"PTM" selects the group of scans from the PTM-peptides to perform the integration.

REMA file named "s2pPTM_OutStats.xls" is created. It contains Zsp for scans from NM and PTM-peptides, and FDRsp for scans only from PTM peptides.

REMAlso a file named "PTM_peptide.xls" is created. It contains Xp and Vp values from PTM-peptides to perform the following peptide-to-protein integration.

```
sanxot.exe -d"full_path/ALL_scan.txt" -r"full_path/s2pALL_Rels.txt" -f -
V"full_path/s2pNM_infoFile.txt" -a"s2pPTM" -o"PTM_peptide.xls" --tags"PTM"


                    REM********** PEPTIDE TO PROTEIN INTEGRATION

REMprepares a joined NM- and PTM-peptide data file

copy "full_path/NM_peptide.xls"+"full_path/PTM_peptide.xls" "full_path/peptide_ALL.xls"


REM--tags"NonMod" selects the NM peptides as the null-hypothesis to calculate the peptide
variance and the protein mean.

REMThe tag is included in the third column of the peptide-to-protein relations file.

REMA file named "protein.xls" is created. It contains protein Xq and Vq values to perform
the following protein-to-category integration.

REMAn additional file is created "p2q_OutStats.xls" containing the Zpq for NM and PTM-
peptides, and the FDRpq for NM peptides only.


sanxot.exe -d"full_path/peptide_ALL.xls" -r"full_path/p2qRels_ALL.xls" -a"p2q" --
tags="NonMod" -o"protein.xls"


                    REM*****************************************

                    REM****** PART 2: SYSTEMS BIOLOGY TRIANGLE *****

                    REM*****************************************

                    REM********* PROTEIN TO CATEGORY INTEGRATION****



REMinitial protein>category integration

REMThe output file "q2c_OutStats.xls" contains Zqc and FDRqc.

sanxot.exe -d"full_path/protein.xls" -r"full_path/q2cRels.txt" -a"q2c_inouts"


REMThe output file "removingOutliers_q2c_tagged.xls" file is created by tagging as "out"
the protein outliers withinh each category at 1% FDRqc in the q2cRels.txt file.

sanxotsieve.exe -d"full_path/protein.xls" -r"full_path/q2cRels.txt" -f0.01 -
V"full_path/q2c_inouts_infoFile.txt" -a"removingOutliers_q2c" --tags="!out"


REMreintegrate protein>category without outliers

sanxot.exe -d"full_path/protein.xls" -r"full_path/removingOutliers_q2c_tagged.xls" -
a"q2c_NoOuts" -o"category.xls" --tags="!out" -f -V"full_path/q2c_inOuts_infoFile.txt"


                    REM********* CATEGORY TO ALL INTEGRATION
```

```
REM-C forces integration to all, obviating the use of a (category-to-all) relations table.
REMThe output file "c2A_OutStats.xls" contains Zca and FDRca.
sanxot.exe -d"full_path/category.xls" -C -v0 -f -a"c2A"

                    REM********* PROTEIN TO ALL INTEGRATION

REMThe output file "q2A_OutStats.xls" contains Zqa and FDRqa
sanxot.exe -d"full_path/protein.xls" -C -V"full_path/q2c_inouts_infoFile.txt" -f -a"q2A"

echo on
```

2.3.1. Trilogy

Trilogy V0.03 is a program made in the Jesus Vazquez Cardiovascular Proteomics Lab at Centro Nacional de Investigaciones Cardiovasculares, used to tag peptides in relations files where non-modified peptides are used as reference in PTM analysis. Trilogy can be downloaded from the link; the program is also supported with the example dataset: https://www.cnic.es/wiki/proteomica/index.php/Trilogy

Input information

```
--version                   show program's version number and exit
 -h,          --help
                            Show this help message and exit
 -a,          --analysis=ANALYSIS
                            Use a prefix for the output files. If this is not
                            provided, then no prefix will be garnered.
 -p,          --place=OUTPUTFOLDER
                            Provide the output folder location, If this not provided
                            then output files will be created in the same location
                            as input file.
 -i,          --inputFile=INPUTFILE
                            Enter the path for inputFile, any tab-separated file can
                            be used as an input. Example: SHIFTS output or Aljamia
                            output
 -c,          --peptideHeader=PEPTIDECOLUMN
                            Identifier for the peptide coloumn
 -C,          --proteinHeader=PROTEINCOLUMN
                            Identifier for the protein/Fasta Description column
 -s,          --ScanHeader=SCANCOLUMN
```

```
                        Identifier for the scan column

-f,         --RawfileHeader=RAWCOLUMN

                        Identifier for the raw filename coloumn

-n,         --NonmodMass=NMAPEXMASS

                        This parameter allows to TAG non-modified peptides.
Tagging can be done in a very flexible manner by using different Tagging
options in combination. Some default tagging options are also included
Check for example option below 1) [-0.000232]|[1.001701]&!D=NonMod.
Multiple things can be considered to a Null-hypothesis in eg.1, [-
0.000232] and [1.001701] both areconsideredas NH, seprated by a |. In
addition, additional conditionscan be used as [1.001701]&!D,
means,[1.001701] shouldbe in peptide but D should notbe. anythying
after = will used as tag for NH

-t,          --PTMtags=PTMTAGS

                        This parameter allows to TAG modified peptides. Tagging
be done in a very flexible manner by using different Tagging options
in combination. Some default tagging options are also included example
1) [15.994862] =Oxidation, [0.010924] =c13|PTMs, in this eg. Multiple
things are considered as modification and all of them have different
label, however for the things, which are not in input, can also be
tagged as modification by using
```

2.4.PtmSticker

PtmSticker is a program made in the Jesus Vazquez Cardiovascular Proteomics Lab at Centro Nacional de Investigaciones Cardiovasculares, used to facilitate the semi-supervised annotation of PTMs.

Input information

```
-h,     --help          show this help message and exit
-i,     --pathforInputfile = INPUTF
                        Enter the path to input file
-d,     --UniModDatabaseFile = DATABASE
                        Enter the path for unimod file
-a,     --Artefectfile = ARTEFACT
                        Enter the path for artefact file
-p,     --peptideColumn = PEPCOL
                        Enter integet coloumn mnumber
-E,     --ErrorExp = EXPERIMENTALERROR
                        Enter the Error value you expect
-e,     --ErrorCombi = COMBINATIONERROR
                        Enter the Error value for mass combination
-O,     --OutPutname=OUTNAME
                        Enter the output file name
```

3. Publication related to PhD work

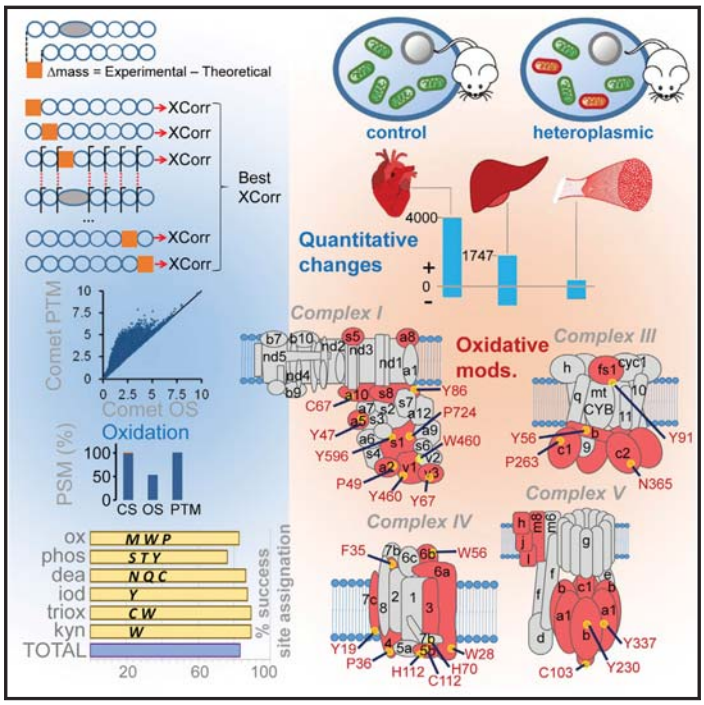
3.1. Comprehensive Quantification of the Modified Proteome Revels Oxidative Heart Damage in Mitochondrial Heteroplasmy

Resource

Cell Reports

Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy

Graphical Abstract



Authors

Navratan Bagwan, Elena Bonzon-Kulichenko, Enrique Calvo, ..., Ana Latorre-Pellicer, José Antonio Enríquez, Jesús Vázquez

Correspondence

jesus.vazquez@cnic.es

In Brief

Bagwan et al. present a suite of algorithms for the unbiased identification and quantification of post-translational modifications and their site of modification by mass spectrometry. They illustrate its utility by showing that mitochondrial heteroplasmy in mice affects mainly the heart, inducing oxidative damage to proteins of the oxidative phosphorylation system.

Highlights

- Algorithms for comprehensive identification of protein modifications by mass spectrometry
- Modified site is located with 85% accuracy
- Integrates quantitative analysis of the proteome and the modified peptidome
- mtDNA heteroplasmy causes oxidative damage in heart OXPHOS proteins

Bagwan et al., 2018, Cell Reports *23*, 3685–3697
June 19, 2018
https://doi.org/10.1016/j.celrep.2018.05.080

CellPress

Cell Reports
Resource

Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy

Navratan Bagwan,[1,5] Elena Bonzon-Kulichenko,[1,2,5] Enrique Calvo,[1,2,5] Ana Victoria Lechuga-Vieco,[1,3] Spiros Michalakopoulos,[1] Marco Trevisan-Herraz,[1,2] Iakes Ezkurdia,[1,2] José Manuel Rodríguez,[1] Ricardo Magni,[1] Ana Latorre-Pellicer,[1] José Antonio Enríquez,[1,4] and Jesús Vázquez[1,2,6,*]

[1]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28049 Madrid, Spain
[2]CIBER Cardiovascular Diseases (CIBERCV), Madrid, Spain
[3]CIBERES: C/ Melchor Fernández-Almagro 3, 28029 Madrid, Spain
[4]CIBERFES: C/ Melchor Fernández-Almagro 3, 28029 Madrid, Spain
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: jesus.vazquez@cnic.es
https://doi.org/10.1016/j.celrep.2018.05.080

SUMMARY

Post-translational modifications hugely increase the functional diversity of proteomes. Recent algorithms based on ultratolerant database searching are forging a path to unbiased analysis of peptide modifications by shotgun mass spectrometry. However, these approaches identify only one-half of the modified forms potentially detectable and do not map the modified residue. Moreover, tools for the quantitative analysis of peptide modifications are currently lacking. Here, we present a suite of algorithms that allows comprehensive identification of detectable modifications, pinpoints the modified residues, and enables their quantitative analysis through an integrated statistical model. These developments were used to characterize the impact of mitochondrial heteroplasmy on the proteome and on the modified peptidome in several tissues from 12-week-old mice. Our results reveal that heteroplasmy mainly affects cardiac tissue, inducing oxidative damage to proteins of the oxidative phosphorylation system, and provide a molecular mechanism explaining the structural and functional alterations produced in heart mitochondria.

INTRODUCTION

Shotgun mass spectrometry (MS)-based proteomics (Link et al., 1999) has become a powerful tool for biotechnological and biomedical research. Advances in speed and sensitivity allow the generation of millions of spectra per experiment, but only a minority of these spectra can be mapped to proteins (Griss et al., 2016; Skinner and Kelleher, 2015). A large proportion of unassigned spectra are thought to arise from peptides containing sequence variants or unknown chemical and post-translational modifications (PTMs) (Griss et al., 2016), and their characterization is one of the most interesting and challenging goals in proteomics. A number of computational methods have been proposed for the detection of these unmatched peptides (Bern et al., 2007; Chen et al., 2009; Griss et al., 2016; Kim and Pevzner, 2014; Ma and Lam, 2014; Shortreed et al., 2015). Recently, an "open search" (OS) strategy, where precursor mass tolerances of hundreds of daltons were used with a conventional search engine, was reported to identify modified peptides at an unprecedented scale (Skinner and Kelleher, 2015). Another report demonstrated that OS can be performed at orders-of-magnitude faster speeds using a fragmentation-ion indexing algorithm (MSFragger) (Kong et al., 2017). These two methods may have a considerable impact on the field, opening the way to true hypothesis-free analysis of PTMs by MS; however, OS algorithms still rely on the chance that the modification leaves enough unaffected fragment ions for matching by the search engine (Figure 1A). OS strategies can therefore identify only approximately one-half of the modified peptides detectable by conventional "closed" searches (CS) (Chick et al., 2015). Moreover, existing OS approaches cannot directly identify the modification site. A further important consideration is that OS methods have not previously been used to quantify PTM alterations, and a general statistical model for the analysis of data of this kind is currently lacking.

Here, we present a suite of bioinformatics tools designed to overcome these limitations. Our tools double the coverage attained by existing algorithms, enabling the generation of comprehensive peptide maps that include practically all of the modifications potentially detectable by MS and CS of the database. Our approach also allows accurate location of the modified residues and quantitative analysis of PTMs in the context of proteome-wide studies. We demonstrate the performance of our tools by performing a comprehensive, tissue-specific characterization of PTMs induced by mitochondrial heteroplasmy in a mouse model. Heteroplasmy has recently attracted the attention of the biomedical community because it can be produced during mitochondrial replacement therapies aimed at preventing transmission of pathogenic mutations in mitochondrial DNA (Craven et al., 2010) or at treating infertility (Wolf

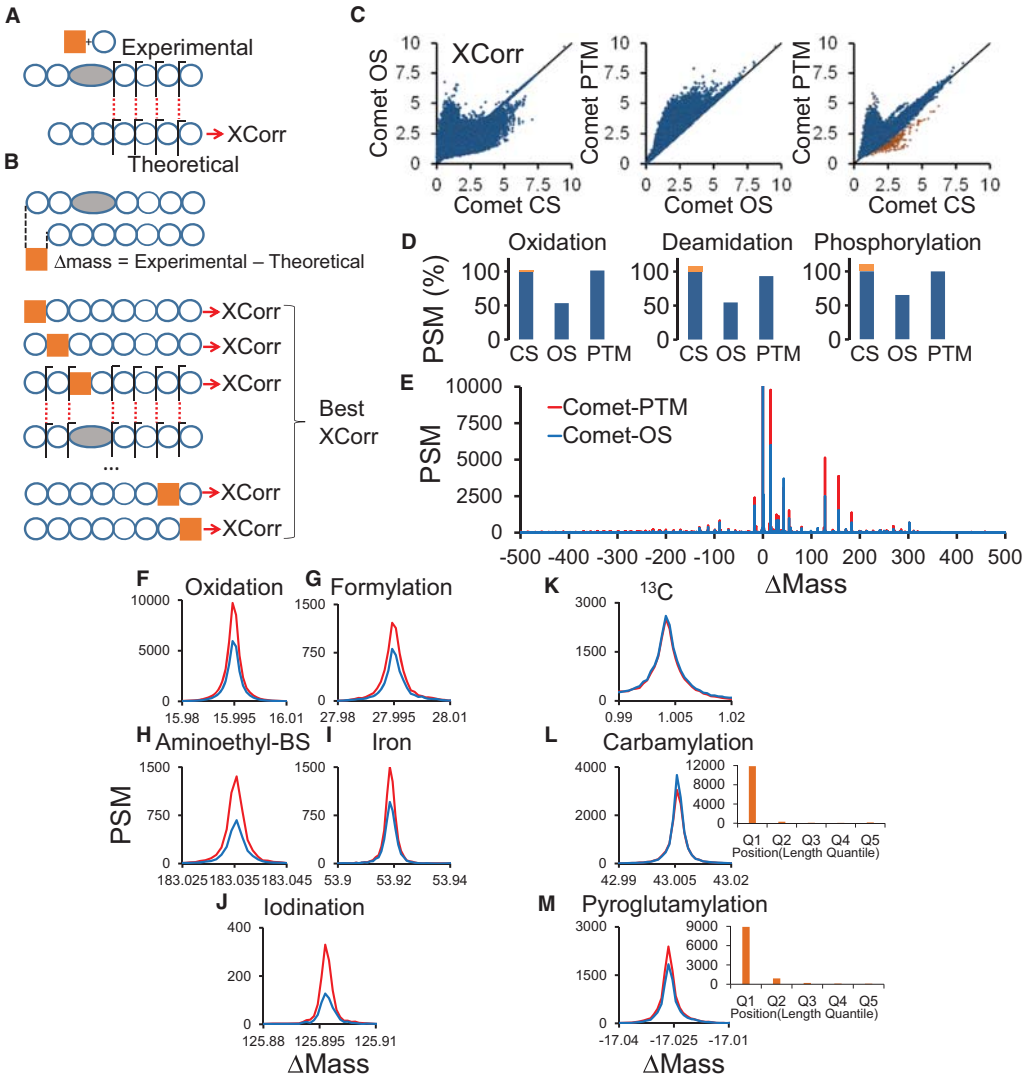


Figure 1. Overview and Peptide Identification Performance of Comet-PTM

(A) Conventional OS methods can identify modified peptides from MS/MS spectra, but only the fragments unaffected by the modification (orange square) are matched; this effect diminishes the score assigned to modified peptides, decreasing identification performance.

(B) Comet-PTM first calculates the difference between the mass of the candidate and the mass of the precursor ion detected by the MS (Δmass). ΔMass is then iteratively added to each amino acid in the peptide sequence, and the position that yields the best score is selected as the correct match.

(C) The HeLa dataset from the original OS article (Chick et al., 2015) was searched using Comet in closed search (CS) mode with three variable modifications (Met oxidation, Asn and Gln deamidation, and Ser and Thr phosphorylation), with Comet in OS mode, or with Comet-PTM (500-Da tolerance in the latter two cases). The scores obtained for the same spectra in the different searching conditions are compared. Yellow points in the rightmost graph are PSMs that match peptides with more than one modification in CS mode.

(D) Identification performance of OS and Comet-PTM relative to CS in the populations of peptides modified by oxidation, deamidation, or phosphorylation. The number of PSMs was obtained after filtering by the score threshold corresponding to 1% FDR in the CS. PSMs matching peptides with more than one modification in CS are indicated by orange bars.

(E) Frequency distribution of PSMs obtained by OS and Comet-PTM as a function of Δmass.

(F–M) Details of the frequency distribution around oxidation (F), formylation (G), aminoethylbenzenesulfonylation (H), iron (I), iodination (J), ^{13}C (K), carbamylation (L), and pyroglutamylation (M). The charts to the right of (L) and (M) show the distribution of PSMs as a function of the quantile position of the modification in the peptide sequence assigned by Comet-PTM.

See also Figure S1.

et al., 2015). Mitochondrial heteroplasmy in mice can be genetically unstable and produce adverse physiological effects (Sharpley et al., 2012). The exact molecular mechanisms that produce the pathological effects are unclear and the potential health risk produced by heteroplasmy in the offspring is debated. Our results show that heteroplasmy between nonpathological mitochondrial DNA variants induces an array of oxidative modifications in the heart that predominantly affect proteins of the oxidative phosphorylation system. The tools presented here thus improve our ability to interpret the totality of information present in MS/MS datasets and provide proteome-wide perspectives for systems biology analysis in high-throughput proteomics.

RESULTS

Comet-PTM Enables Comprehensive Identification of Peptide Modifications

We developed Comet-PTM, an improved OS engine that takes into account the mass shift produced by the modification in the fragmentation series, producing the same score as a CS search using the same mass increment as a variable modification (Figure 1B). To test the performance of Comet-PTM, the results were compared with those obtained with OS and with CS using three common variable modifications. As expected, OS assigned higher scores to a large population of peptide-spectrum matches (PSMs) containing modifications not included in the CS list of variable modifications (Figure 1C, left). However, CS assigned a higher score than OS to a large population of PSMs with modifications included in the CS list, because the modifications affected the matching of fragment ions, decreasing the OS-assigned score (Figure 1C, left). This effect reduced the identification performance of OS to around 50% of the PTMs identified by CS (Figure 1D), confirming previous results (Chick et al., 2015). In contrast, scores obtained with Comet-PTM matched or exceeded OS (Figure 1C, center). In addition, Comet-PTM identified the same Δmass peaks as OS (Figure 1E); however, most of these peaks contained about 2-fold more PSMs (Figures 1E–1J), reflecting the superior performance. As expected, this effect was not observed when the modification did not affect the fragmentation series. Thus, the ^{13}C peak, produced by errors in assignment of the monoisotopic precursor mass, was observed with exactly the same number of PSMs (Figure 1K). The number of PSMs was also similar for modifications in N-terminal position of the peptide, which only affect b-series and have a negligible effect on identification by higher-energy collisional dissociation (HCD) fragmentation (Figures 1L and 1M).

Comet-PTM produced scores similar to or higher than those obtained with CS (Figure 1C, right), except for a small population of peptides for which CS found two or more modifications (Figure 1C, orange dots). This effect was not due to differences between CS and Comet-PTM scores (Figures S1A and S1B). Therefore, and unlike OS, Comet-PTM had a similar identification performance to that of CS for the preselected modifications (Figure 1D). In several instances, Comet-PTM correctly located an oxidation on Trp, Pro, or Tyr that CS wrongly assigned to oxidation on Met, the predefined variable modification residue in CS (Figures S1C–S1E). This finding shows that conventional PTM searches using a variable modification on selected amino acids can force false assignation of the modified residue. Taken together, these data show that the Comet-PTM OS engine efficiently resolves the modification mass shift in the fragmentation series, doubling the number of modified peptides identified by conventional OS and matching the identification performance of targeted CS.

Comet-PTM Detects the Location of Modifications in the Peptide Sequence

Peptides were identified after Comet-PTM searches with SHIFTS, an algorithm that detects the peaks in the Δmass distribution and controls the peptide false-discovery rate (FDR) through a conservative, three-layered approach (Figures 2A–2C; see also STAR Methods). From the output of a search against a concatenated target-decoy database, SHIFTS calculates a global score threshold to control FDR in the population of PSMs with Δmass greater than −56 Da (Figure 2A). A local score threshold is also defined to control FDR separately within each of the ~1-Da bins in the Δmass distribution (Figure 2B). Finally, a peak score threshold is calculated to control FDR separately within each peak detected in the Δmass distribution (Figure 2C). A PSM is considered to be correctly identified when its score is above the global and peak thresholds; when a PSM does not form part of a peak, instead of the peak threshold the local threshold is used. This conservative approach allows full control of FDR, avoiding any bias due to the specific behavior of certain kinds of peptide modifications that may be more prone to match decoy sequences.

An advantage of Comet-PTM over existing OS approaches is that it automatically assigns the modification to the residue in the peptide sequence that produces the best score and therefore is the modified position that best explains the fragmentation data. Assigning modifications to specific residues is considered a much less reliable process than identifying peptides, partly because there is frequently insufficient information to determine the exact modified residue (Chalkley et al., 2008). To estimate the accuracy of Comet-PTM assignation of a modification to the correct site, we filtered correct identifications using the conservative method described above with a 1% FDR threshold and then calculated the fraction of PSMs in which the mass shift was located in amino acids predicted to harbor well-known modifications. Oxidation was located to Met, Trp, or Pro in 82% of cases, and deamidation to Asn, Gln, or carbamidomethylated Cys in 86% of cases (Figure 2D). Considering six well-known modifications, the overall accuracy of Comet-PTM was estimated at ~85%. To test more specifically the performance of Comet-PTM for the correct assignation of phosphorylation sites, we analyzed a dataset obtained from synthetic phosphorylated peptides (Ferries et al., 2017). Comet-PTM assigned the correct position with 81% accuracy, and in 94% of the cases the modification was in the correct or in an adjacent position (Table S1).

We then analyzed which amino acids were assigned to known modifications with a higher frequency than expected by chance. Among the most frequent, we only found well-known modifications like oxidized Met and Trp, deamidated Asn, Gln and carbamidomethylated Cys, phosphorylated Ser, trioxidized Cys and

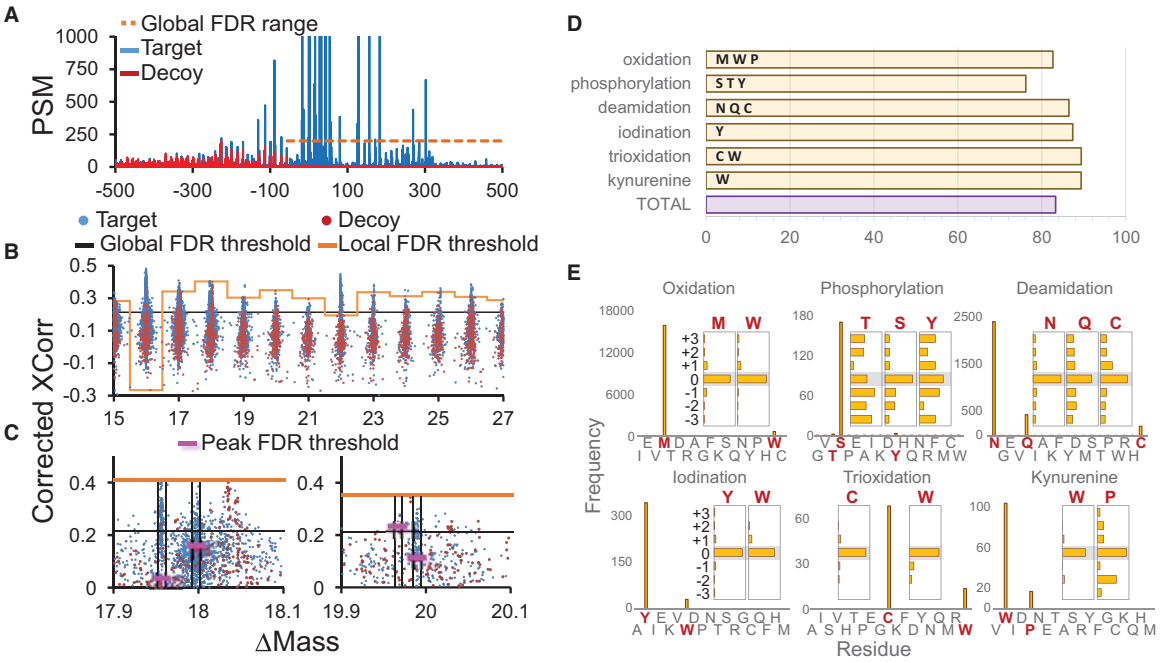


Figure 2. Identification of Modified Peptides and Location of the Modified Site

(A) ΔMass distribution of target and decoy PSMs and range of application of the global score threshold. The vertical scale is enlarged to show the distribution of decoy PSMs.

(B) Local score thresholds applied in each ~1-Da bin in the Δmass distribution.

(C) Peak score thresholds applied to the peaks detected by SHIFTS in the Δmass distribution.

(D) Percentage of successful assignment of the indicated modification to the indicated residues.

(E) Inset horizontal bar graphs plot the frequency distributions of PSMs assigned by Comet-PTM to the indicated amino acids and modifications. Vertical bar graphs: frequency of PSMs assigned to the indicated amino acids after subtracting the average frequency of the three previous and the three subsequent positions. Residues that accumulate counts in the histogram at position 0 are labeled in red.

See also Figures S2 and S3.

Trp, and kynurenine-modified Trp (Figure 2E). Some Pro modifications were also found with the Δmass of kynurenine-modified Trp (Figure 2E), but all of them corresponded to homologous peptides that contained the Pro>Thr substitution, which has the same Δmass (Figure S2), and that were assigned the same score. Finally, Trp iodination was also clearly detected above background (Figure 2E). This modification was unexpected because, although Trp halogenases have been described in bacteria (van Pée and Patallo, 2006), they are not present in mammals, and Trp iodination has not been described before as a chemical artifact. The majority of peptides containing iodinated Trp contained neither Tyr nor His, the two amino acids known to react with iodine. A careful inspection of their fragment spectra confirmed the presence of y-fragments that could only be explained by Trp being the modified residue (Figure S3). In addition, the only modification described in Unimod that matched the observed mass shift corresponded to iodination. From all of these data, we concluded that Trp was iodinated in these peptides, most probably as a side-chain reaction of the iodine produced from the Cys-alkylating reagent iodoacetamide. This finding further highlights the accuracy with which Comet-PTM locates the site of modification and suggests that this OS engine may be useful for detecting novel peptide modifications in specific amino acids.

A Single Integrated Statistical Framework Allows Quantification of the Proteome and of the Modified Peptidome

For the quantitative analysis of modified peptides, we developed an algorithm based on a previously proposed systematic workflow (García-Marqués et al., 2016; Navarro et al., 2014) (Figure 3A). The algorithm includes a peptide-to-protein integration step that quantifies protein values from the unmodified peptide forms and then computes the standardized log2-ratio of the modified peptides with respect to these protein values (Z_{pq})(Figure 3B). This makes it possible to detect modified peptides whose behaviors deviate significantly from those of the unmodified peptides from the same protein. Note that the algorithm calculates peptide abundances relative to the parent protein, not to the mean of all the peptides in the sample, so that changes at the peptide level are unaffected by changes in protein abundance. When applied to a biological model, the Z_{pq} distributions of the

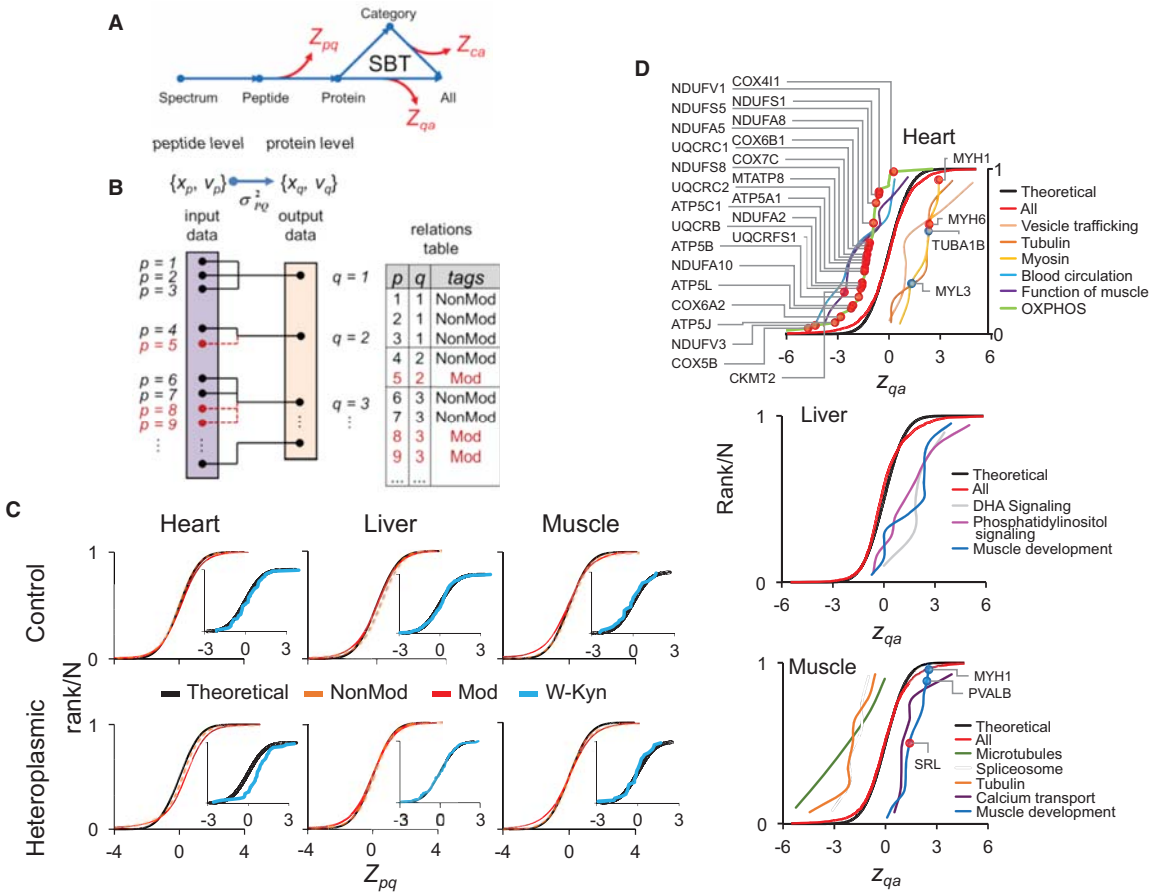


Figure 3. Quantitative Analysis of Alterations Produced by Heteroplasmy in the Modified Peptidome and in the Proteome

(A) Scheme of the integrative workflow used to quantify modified peptides, proteins or categories. Each arrow represents a step performed with the generic integration algorithm (García-Marqués et al., 2016). Standardized $\log_2$-ratio values at the peptide level (Z_{pq}) are obtained from the peptide-to-protein integration. The algorithm provides corrected peptide values by taking account of the corresponding protein changes. The same workflow is used for protein quantification and systems biology analysis of coordinated protein responses with the SBT model (García-Marqués et al., 2016).

(B) Scheme of the generic integration algorithm (GIA) (García-Marqués et al., 2016) adapted for the analysis of modified peptides in the peptide-to-protein integration step. The relations table include tags that can be used to indicate that modified peptides are excluded from the computation of protein averages.

(C) Distributions of Z_{pq} values for nonmodified (Non-Mod) and modified peptides (Mod). These data come from the analysis of mouse tissues in a model of mitochondrial heteroplasmy; each graph corresponds to a tissue from an individual animal. The inset plots show the Z_{pq} distributions for peptides containing the kynurenin Trp modification. Note that the distribution is displaced to the right in the hearts of heteroplasmic mice, indicating increased abundance. For further details, see STAR Methods.

(D) Characterization of coordinated protein alterations produced by heteroplasmy in heart, liver, and muscle using the SBT model. Z_{qa} values are standardized $\log_2$-ratio averages of proteins from heteroplasmic mice relative to controls. Proteins are grouped into representative functional categories with a statistically significant change. The upper panel presents proteins in the OXPHOS category that are decreased in the heteroplasmic heart and that contain heteroplasmy-induced oxidative modifications (Figure 6).

See also Table S2.

unmodified and modified peptide forms precisely followed the expected null hypothesis distribution in different tissues from two mouse strains (Figure 3C). These results convincingly demonstrate that the quantitative behavior of the entire population of modified peptides can be modeled accurately, providing a robust statistical framework within which to analyze abundance changes in high-throughput experiments. This allowed detection of significant changes in certain modified peptides against the null hypothesis in specific situations (Figure 3C, lower left).

The same workflow used for peptide quantitation allowed the analysis of protein abundance changes and the characterization of functional categories that were affected by the coordinated action of proteins, using the systems biology triangle (SBT) model (García-Marqués et al., 2016) (Figure 3A). This statistical

design thus allowed a full description of the alterations in the modified peptidome in the global context of changes in the proteome.

Heteroplasmy Produces Protein Alterations Consistent with a Mitochondrial Dysfunction in the Heart

We used Comet-PTM and associated tools to study the molecular impact of mitochondrial heteroplasmy on the proteome and the modified peptidome in heart, liver, and skeletal muscle from 12-week-old mice. This age was considered the most appropriate to study molecular alterations before more extensive damage was produced as a consequence of heteroplasmy. SBT model analysis of the quantitative protein data from the heart revealed a coordinated decrease of proteins related to muscle function and the oxidative phosphorylation system (OXPHOS) and a coordinated increase of tubulins, myosins, and proteins involved in vesicle trafficking. The muscle-function proteins included KCRS (mitochondrial creatine kinase) and VDAC3, which are key suppliers of phosphocreatine to the sarcomere, and SGCG and KGP1, which are essential for sarcomere contraction (Table S2). These alterations in heteroplasmic mice are highly consistent with the decreased ATP synthesis and the abnormal increase in phosphocreatine/ATP ratio in the heart and with the increase in plasma creatine kinase we have observed in this animal model (unpublished data), and provide evidence that heteroplasmic cardiac mitochondria have a compromised ability to supply energy to the sarcomere. Among the increased vesicle trafficking proteins, STXB3 and VAMP2 are implicated in the insulin-dependent movement of GLUT4 from inner vesicles to the plasma membrane. This finding also agrees with a higher glucose uptake we detected in heteroplasmic hearts, pointing to a shift toward glycolytic metabolism in order to compensate the reduced phosphocreatine supply from the OXPHOS. The coordinated decrease in blood proteins and the increase in heart cytoskeletal proteins in heteroplasmic animals suggest a homeostatic effort to maintain cardiac cell structure.

Heteroplasmy did not affect OXPHOS protein levels in liver but did produce a coordinated increase of docosahexaenoic acid (DHA) signaling pathway proteins (Figure 3D), including AKT2, BID, and B2CL1 (Table S2). AKT2 plays an essential role in the progression of the inflammatory response and apoptosis (López-Carballo et al., 2002); consistent with this finding, we detected signs of inflammation in the liver of heteroplasmic mice with aging (unpublished data). Heteroplasmic liver also showed coordinated increases in a group of myosins and phosphatidylinositol pathway proteins, probably reflecting a homeostatic response. These data thus indicate that, at 12 weeks of age, heteroplasmy produces a decrease in OXPHOS proteins in the heart, probably reflecting mitochondrial dysfunction, but the same effect is not seen in liver.

In heteroplasmic muscle, we detected a coordinated decrease of proteins regulating cell shape, such as tubulins and microtubule proteins (Figure 3D). On the other hand, the thick-filament myosins, as well as calcium transporting proteins, were increased, suggesting the induction of compensatory mechanisms. However, we found no evidence of altered mitochondrial protein levels in muscle at this age.

Heteroplasmy Mainly Produces Oxidative Modifications of OXPHOS Proteins in Heart

To study the impact of heteroplasmy on the modified peptidome, we first compared the Δmass distribution of peptides searched with Comet-PTM and identified with SHIFTS in the three tissues. The distribution of the most abundant peaks was very similar in liver, heart, and muscle (Figure S4). In addition to the most intense peak corresponding to unmodified peptides, we found peaks corresponding to oxidized and deamidated peptides, and dioxidized peptides were also frequent. Other frequent modifications were artifacts produced by TMT reagents and iodoacetamide treatments, sodium adducts, and missed cleavages. Some tissue-specific modifications were also detected, mostly in the region from 50 to 100 Da. Among these, phosphorylated peptides were more abundant in muscle and 90% of these peptides belonged to proteins implicated in muscle contractility, consistent with the regulatory role of phosphorylation in the sarcomere. Within each Δmass peak, we analyzed the distribution of the modified amino acid residues, which were assigned to frequent modifications using the unbiased approach followed in Figure 2E. Confirming the results described above, amino acids known to harbor specific modifications were mostly located correctly (Figure S5). Interestingly, the relative proportion of modified sites was remarkably preserved across the three tissues, with the exception of tyrosine iodination, which was prominent in heart but less frequent in liver and skeletal muscle (Figure S5). This analysis also detected the increased frequency of phosphorylation in muscle (Figure S5). We estimated that about one-quarter of the total peptidome in the three tissues had known PTMs, mostly oxidation and deamidation; less than 10% of modifications were chemical artifacts (Figure 4A).

Second, we analyzed the quantitative data with the integrative statistical algorithm (Figure 3A) to determine the pattern of modifications that are specifically affected by heteroplasmy in each tissue. Heteroplasmy induced marked tissue-specific modifications, most of them concentrated in heart and barely detectable in skeletal muscle (Figure 4B; Table S3). The vast majority of increased modifications in heart were oxidative, affecting mostly Tyr, Trp, Pro, and Phe and to a lesser extent Asn, Cys, and Asp (Figure 4C). Heteroplasmy also induced phosphorylation and methylation in the heart. The modification pattern was similar in liver, but the effect of heteroplasmy in this tissue was significantly less pronounced.

Functional enrichment analysis revealed that heteroplasmy-induced modifications in mouse heart were mostly located in mitochondrial proteins, particularly those located in the inner mitochondrial membrane; moreover, most of the modified proteins were OXPHOS proteins, which were enriched with multiple oxidative modifications (Figure 5A). Tricarboxylic acid (TCA) cycle proteins and those related to the oxidoreductase complex were enriched in several modifications. Deamidation affected several mitochondria-related categories, and phosphorylation affected proteins related to contractility. Heteroplasmy-induced modifications affected considerably fewer categories in liver (Figure 5B) and were not significantly enriched in muscle (Figure 5C), where only a few categories were decreased, containing low numbers of proteins. Deamidated peptides were found in proteins related to mitochondria, NAD binding, and extracellular

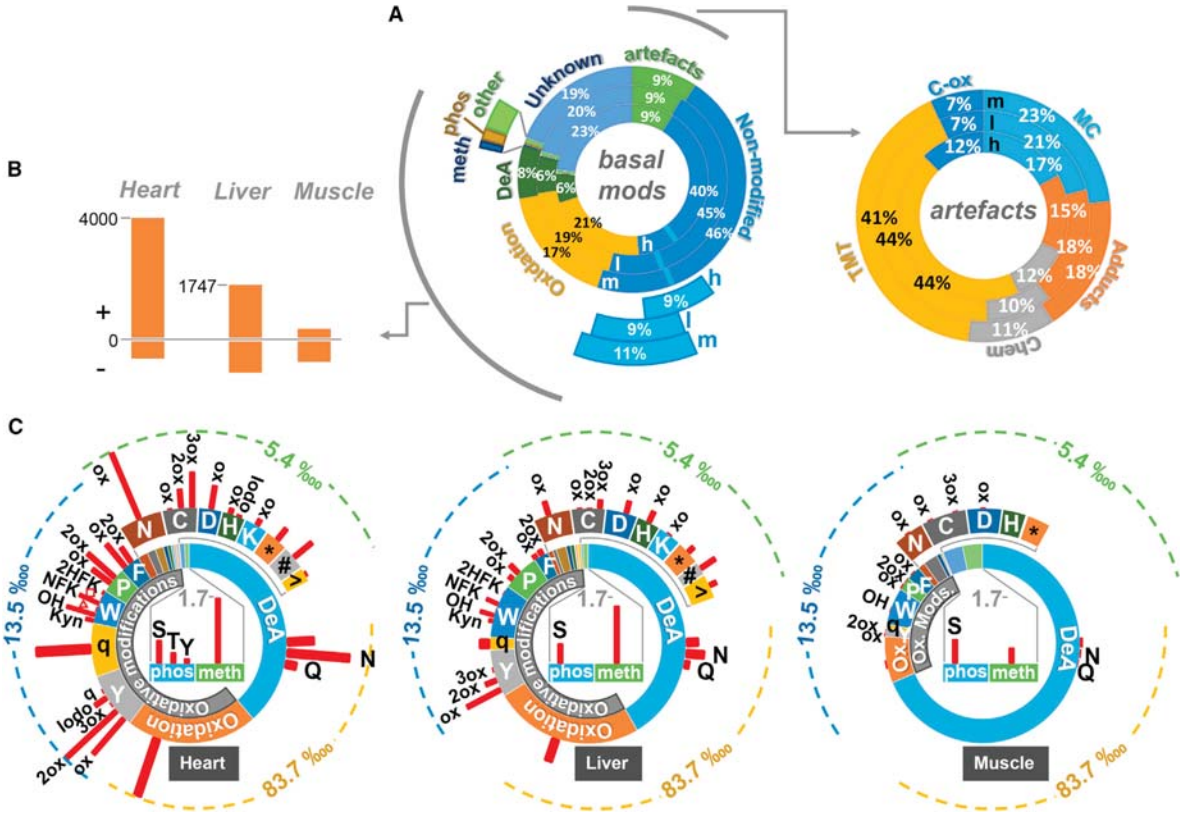


Figure 4. Alterations Produced by Heteroplasmy in the Modified Peptidome in Different Mouse Tissues

(A) Complete map of the basal peptidome, expressed as a percentage of the total number of peptides identified in each tissue (circular bar graphs: h, heart; l, liver; m, muscle). For simplicity, peptides oxidized on Met (light blue) are included in the group of nonmodified peptides. The plot to the right shows proportions of artifacts produced by sample manipulation.

(B) Number of PTM-containing peptides significantly increased (+) or decreased (−) in heteroplasmic mouse tissues (n = 4) relative to controls (n = 3).

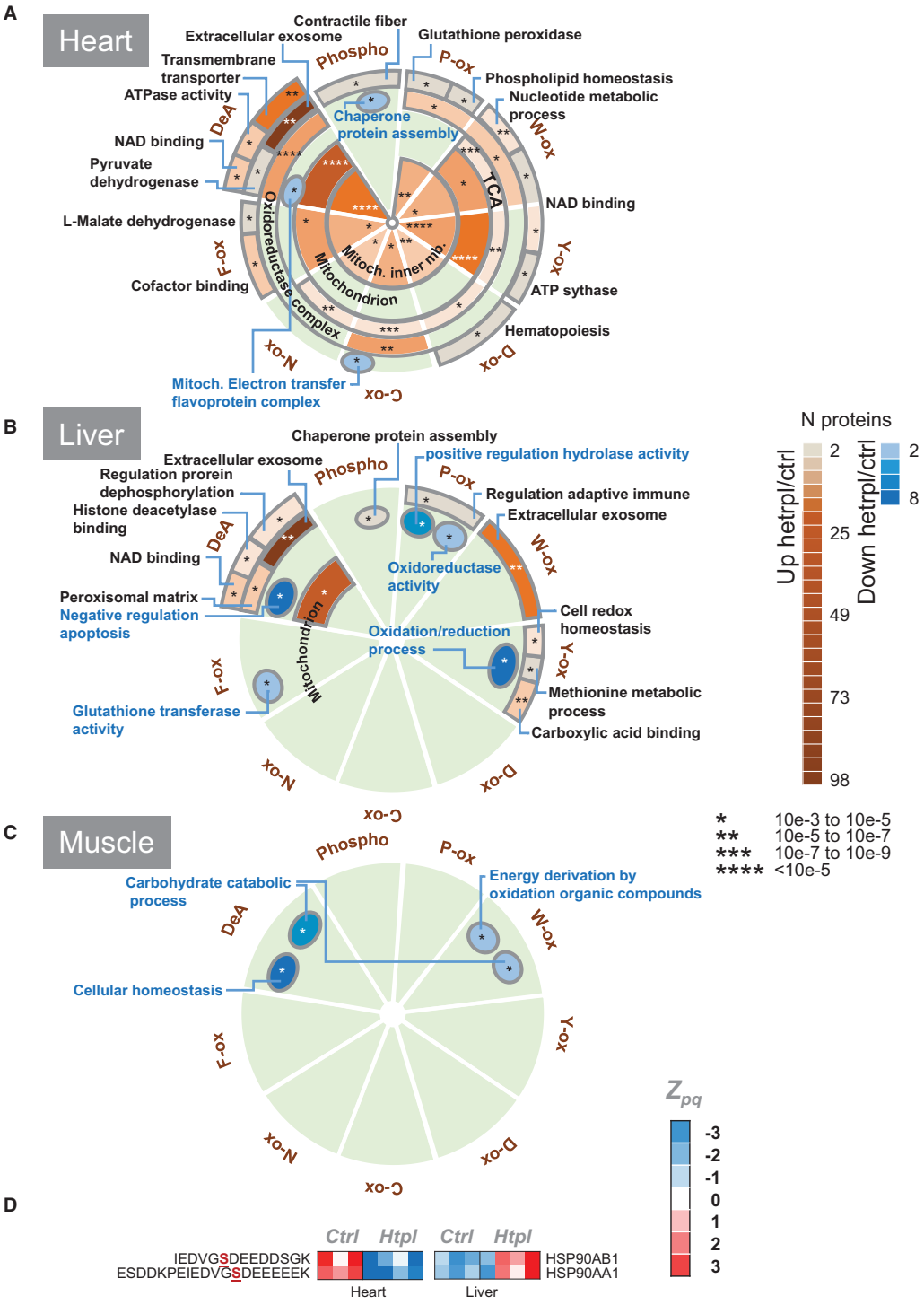
(C) Statistically significant PTM increases in heteroplasmic mouse tissues according to the type of modification and the modified residue. The circular inner bars show the peptide proportion for each modification in the basal state, and the radial red bars represent the proportion of peptides of each kind (in parts per 10,000) that are increased in heteroplasmic tissues. For the sake of clarity, three different scales are used depending on the type of modification. DeA, deamidation; meth, methylation; phos, phosphorylation; MC, missed cleavages; Chem, chemical derivatives produced by sample preparation; TMT, extra addition of the isobaric labeling reagent; Adducts, Na, K, and ammonia adducts. Uppercase letters indicate the single-letter amino acid code. ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; red triangle, Trp to oxolactone; red square, Trp to quinone; q, quinone, Tyr to quinone; Iodo, iodination; asterisk, Met to homocysteic acid; hashtag, Met to homoserine.

See also Figures S4 and S5 and Table S3.

exosomes in both heart and liver, suggesting that these modifications do not contribute to tissue-specific phenotypic differences between heteroplasmic and control mice.

Interestingly, heteroplasmy decreased phosphorylations in proteins related to chaperone activity in the heart but had the opposite effect in liver (Figures 5A and 5B). This category contains the two HSP90 isoforms (HSP90AA1 and HSP90AB1), which harbor PTM sites involved in the regulation of the chaperone activity (Mollapour and Neckers, 2012). These isoforms are known to be phosphorylated at Ser263 and Ser255, respectively (Hu et al., 2015; Wang et al., 2017), within the CK2 phosphorylation motif (S-X-X-E), which is crucial for protein activity. In particular, phosphorylation at Ser255 is required for the activation of mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) pathway. Comet-PTM and SHIFTS identified these two phosphorylated sites, and heteroplasmy decreased their levels in heart while increasing them in liver (Figure 5D). These findings suggest that heteroplasmy differentially affects chaperone activity and support the roles in cardiac heteroplasmy we have proposed for the endoplasmic reticulum and the mitochondrial unfolded stress response (unpublished data).

Heteroplasmy also altered phosphorylation levels at sites known to regulate protein activity in a tissue-specific manner. Phosphorylation at Ser42 from PTRF, which regulated transcriptional activity in response to metabolic challenges (Huttlin et al.,



(legend on next page)

2010; Liu and Pilch, 2016), was decreased in heart but was increased in muscle, in agreement with the different metabolic status of these tissues (Table S3). Similarly Ser133 from MYG, which regulates contractility (Lundby et al., 2013), was more phosphorylated in heart, in line with the coordinated decrease of proteins related to muscle function (Figure 3D), suggesting alterations in the contractile function of the heart in heteroplasmic animals. In the liver, the active site Ser117 from PGM1, which regulates glycogen biosynthesis (Huttlin et al., 2010; Lee et al., 2014), was less phosphorylated, suggesting lower fuel storage in this tissue (Table S3). These results reinforce the notion that the impact of heteroplasmy is tissue specific. The rest of phosphorylated sites affected by heteroplasmy had unknown regulatory roles, but it should be noted that all of them have been described previously, supporting the accuracy of the algorithms used to detect them. They include Ser181 in PGRMC1 (Jin et al., 2004), Ser442 in SRCA (Dai et al., 2007), Ser475 in JPH1 (Rigbolt et al., 2011), and Ser298 and Ser304 in BIN1 (Ballif et al., 2004; Huttlin et al., 2010) (Table S3).

To compare the alterations in the modified peptidome with the changes observed at the protein level, we mapped the proteins harboring PTMs affected by heteroplasmy onto the protein distribution plots depicting coordinated protein responses (red and blue dots in Figure 3D). Most increased modifications mapped onto the OXPHOS protein category in heart, which was decreased, whereas proteins belonging to other categories were mostly unaffected. This finding suggests that the oxidative damage produced by heteroplasmy in heart OXPHOS proteins induced their degradation, and is consistent with the increase in oxidative stress we have observed in heart slices, embryonic fibroblasts, and adult fibroblasts of heteroplasmic mice (unpublished data).

Most heteroplasmy-induced oxidative modifications in OXPHOS proteins occurred in the four respiratory protein complexes, except for Trp oxidations, which were mostly concentrated in C-IV (Figure 6). The modifications were clustered in groups of proteins located together in the structure of the complexes and predominantly affected proteins oriented toward the mitochondrial matrix, where proteins are more exposed to mitochondrial reactive oxygen species (ROS). In further analysis, we used Consurf to calculate evolutionary conservation profiles for the affected residues (Ashkenazy et al., 2016) and mapped the modifications onto the known three-dimensional (3D) structures of the complexes, when available, or onto their predicted models (Meier and Söding, 2015). We found that 70% of the modified residues were conserved or interacted with conserved regions (Table S4). Moreover, most of the modifications were located on the protein surface, and those that were buried tended to have high conservation scores, suggesting they are important for structural stability or folding (Table S4). These findings suggest that these oxidative modifications compromise OXPHOS protein function. This view is supported by the decreased ATP synthesis capacity we have detected in heteroplasmic heart.

DISCUSSION

Comprehensive analysis of PTMs in mass spectrometry data obtained through shotgun approaches has historically presented a bioinformatics challenge because conventional (or closed) database searches can only identify predefined subsets of modifications. This limitation impedes truly hypothesis-free analysis of modifications, to the point that the proportion of chemical and PTM in biological systems has remained unknown. The open database search strategy (Chick et al., 2015) provided a starting point for resolving this problem, allowing a truly unbiased high-throughput identification of modified peptides without previous knowledge on the nature of these modifications. An improved fragment-ion indexing method (MSFragger) later demonstrated that OS can be performed on a timescale similar to or shorter than that required for CS (Kong et al., 2017). The open-search strategy has the additional advantage that modified and nonmodified peptide forms are identified simultaneously (Kong et al., 2017). These strategies, however, are not comprehensive and provide limited information on the nature of modified residues. The computational approach presented here enables the open-search strategy to match the performance of CS, identifying the vast majority of detectable peptide modifications. Although Comet-PTM is built on the framework of a well-known database searching engine, it is important to note that the underlying concept can be extrapolated to any other searching engine, opening the way to near-complete analysis of the protein-modification landscape in biological systems using tools already used by the proteomics community. In addition, our method provides information on the most probable site containing the modification with an estimated accuracy of ~85%. It should be noted that, in its current state of development, the method might fail if the peptide contains more than one modification. However, the impact of this limitation on the general performance of the algorithm seems small, since the fraction of multiply modified peptides that were not identified was almost negligible. We also introduced a conservative, three-layered approach to control the FDR of peptide identification. Our local FDR calculation is similar to that used by MSFragger (Kong et al., 2017), with the only difference that we used the decoy/target competition strategy to estimate FDR, whereas Kong et al. modeled local score distributions, in the same 1-Da bins, as a mixture of correct and incorrect identifications and calculated the probability of correct identification by the Bayes rule. As recognized by these authors, their approach did not take into account the parts-per-million levels of accuracy produced by high-resolution

Figure 5. Functional Distribution of Proteins Post-translationally Modified by Heteroplasmy

(A–C) Functional enrichment analysis of the proteins containing PTMs that are significantly increased (brown scale) or decreased (blue scale) in heart (A), liver (B), or muscle (C) of heteroplasmic mice in relation to control mice. The analysis was performed separately per different kinds of modification (slices). Color intensity represents the number of proteins in each category. Asterisks indicate statistical significance of category enrichment relative to the list of proteins identified in each tissue. Inner circular rings are used for categories at the cellular component level, whereas outer rings correspond to biological processes.

(D) Heatmap of the Z_{pq} values for two modified proteotypic peptides from the two HSP90 isoforms (HSP90AB1 and HSP90AA1) for all control and heteroplasmic replicates in the heart and in liver. These two isoforms are included in the "chaperone protein assembly" functional category displayed in (A) and (B).

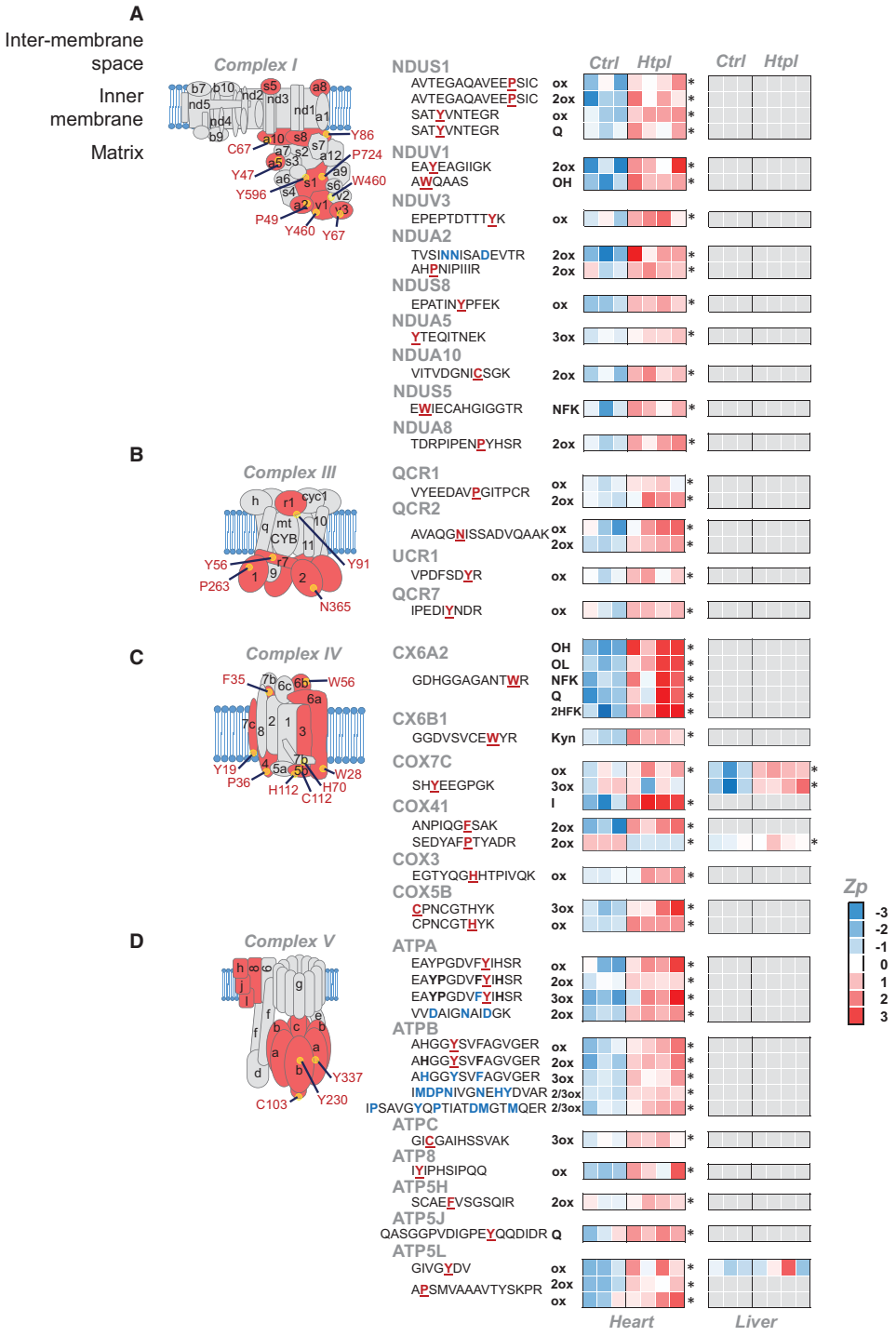


Figure 6. PTMs Induced by Heteroplasmy in OXPHOS Proteins

(A–D) Scheme of heteroplasmy-induced PTMs in proteins of complex I (A), complex IV (B), complex III (C), and complex V (D) from the OXPHOS system. PTM-containing proteins are highlighted in red, with the spatial position of the modified residue indicated. The heatmaps show statistically significant changes in

(legend continued on next page)

instruments (Kong et al., 2017), which we have considered in our model by applying an additional peak FDR estimation. In addition, we found it necessary to introduce a third global FDR filter to avoid identification of low-scoring peptides in local bins or in peaks that, as in the case of oxidation, concentrate high numbers of target PSMs.

The algorithm allowed us to produce peptide maps containing realistic proportions of unmodified peptides, PTMs, and chemical artifacts in mouse tissues, and of amino acid residues subjected to each type of modification. We show that the basal distribution of modifications and modified amino acids is remarkably similar across the three mouse tissues studied, indicating that the algorithm generates reproducible results and that most peptide modifications are of an oxidative nature. Many of the modifications are introduced during sample preparation, including those induced by exposure to Cys-alkylating and amine-directed isobaric labeling reagents; although these agents are thought to be highly specific, in practice they produce secondary reactions with other nucleophiles.

The ability to perform a truly hypothesis-free high-throughput analysis of modifications can contribute to a deeper understanding of the relation between protein chemistry and function; however, the most important biological insight obtainable with shotgun proteomics is how these modifications are altered during normal physiological processes or as a consequence of disease. Detecting these specific alterations allows to concentrate the study on biologically meaningful modifications, obviating the analysis of the abundant artifacts produced by sample preparation. The statistical approach presented here enables quantitative analysis of the modified peptidome within the same framework used to analyze the unmodified proteome; this approach thus allows coherent interpretation of all of the results and opens the way to integrated systems-biology analysis. We demonstrate the performance of our algorithm by simultaneously characterizing, in three tissues of a mouse model, the quantitative impact of heteroplasmy on the proteome and on the modified peptidome. To our knowledge, this is the first study to apply open-search-based approaches together with quantitative MS, allowing comprehensive and unbiased characterization of alterations in the protein modification landscape produced in a biological system.

Our results reveal that the major hallmark of mitochondrial heteroplasmy in 12-week-old animals is oxidative damage to OXPHOS proteins in the heart. This is mainly reflected in the increased levels of oxidative modifications and in the decreased levels of OXPHOS proteins. Our data support the notion that heteroplasmy involving different nonpathological mtDNA variants affects the performance of the OXPHOS system in the heart, resulting in compromised mitochondrial ROS handling that triggers an oxidative stress response and impairs the ability to supply energy to the sarcomere. Heteroplasmy is known to be related to ROS production (Hämäläinen et al., 2013, 2015) and also affects the levels of contractile proteins in muscle. These alterations do not occur in liver, which is able to resolve heteroplasmy (unpublished data). Our data thus provide a molecular mechanism that explains the functional findings we have observed in the heteroplasmic heart, which include a diminished ability of mitochondria to produce ATP, increased glucose uptake, an exacerbated phosphocreatine/ATP ratio, and mitochondrial disruption and structural alterations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse Model of Heteroplasmy
  - Generation of Heteroplasmic Mice
  - Mice Breeding
- METHOD DETAILS
  - Benchmarking MS Dataset
  - Preparation of Protein Extracts
  - Protein Digestion, Peptide Labeling, and Fractionation
  - LC-MS Analysis
  - Database Search
  - Comet-PTM
  - SHIFTS
  - Annotation of Modifications
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Peptide Quantification and Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY
  - Code Access

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.05.080.

ACKNOWLEDGMENTS

We thank Simon Bartlett (CNIC) for English editing. This study was supported by competitive grants from the Spanish Ministry of Economy and Competitiveness (MINECO) (BIO2015-67580-P) through the Carlos III Institute of Health-Fondo de Investigación Sanitaria (PRB2, IPT13/0001-ISCIII-SGEFI/FEDER; ProteoRed), by Fundación La Marato TV3, and by FP7-PEOPLE-2013-ITN "Next-Generation Training in Cardiovascular Research and Innovation–Cardionext." N.B. is a FP7-PEOPLE-2013-ITN-Cardionext Fellow. The CNIC is supported by the MINECO and the Pro-CNIC Foundation, and is a Severo Ochoa Center of Excellence (MINECO Award SEV-2015-0505).

abundance (Z_{pq}) of the indicated peptides in heart and liver from control mice (Ctrl) and heteroplastic mice (Htpl). Gray squares indicate that the modified peptide was not detected in the tissue sample. Modified residues that are unambiguously assigned are underlined bold red letters; tentatively assigned sites are in bold blue letters. Asterisks mark statistically significant changes between control and heteroplasmic samples according to Student's t test ($p < 0.05$). The modification type is indicated in the column to the right of the peptide sequences: ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; OL, Trp to oxolactone; Q, Trp to quinone; I, iodination. The conservation scores of these peptides are listed in Table S4, from which the results of Vseq analysis can also be downloaded.See also Table S4.

OPEN ACCESS
CellPress

AUTHOR CONTRIBUTIONS

N.B., E.B.-K., E.C., and J.V. designed the algorithms. S.M. and M.T.-H. programmed Comet-PTM code. N.B. programed SHIFTS and analyzed Comet-PTM performance. N.B., E.B.-K., and M.T.-H. programmed the generic integration algorithm and performed the quantitative analysis of PTM. A.V.L.-V., A.L.-P., and J.A.E. developed the heteroplasmic mouse model. E.C., A.V.L.-V., and R.M. prepared the tissue extracts. E.C. and R.M. performed the mass-spectrometry analysis. I.E. and J.M.R. performed the bioinformatic analysis of PTM. N.B., E.B.-K., E.C., and J.V. wrote the manuscript. J.V. directed the research.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 8, 2018
Revised: April 13, 2018
Accepted: May 23, 2018
Published: June 19, 2018

REFERENCES

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44* (*W1*), W344–W350.

Ballif, B.A., Villén, J., Beausoleil, S.A., Schwartz, D., and Gygi, S.P. (2004). Phosphoproteomic analysis of the developing mouse brain. Mol. Cell. Proteomics *3*, 1093–1101.

Bern, M., Cai, Y., and Goldberg, D. (2007). Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. Anal. Chem. *79*, 1393–1400.

Bonzon-Kulichenko, E., Garcia-Marques, F., Trevisan-Herraz, M., and Vázquez, J. (2015). Revisiting peptide identification by high-accuracy mass spectrometry: problems associated with the use of narrow mass precursor windows. J. Proteome Res. *14*, 700–710.

Chalkley, R.J., Baker, P.R., Medzihradszky, K.F., Lynn, A.J., and Burlingame, A.L. (2008). In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol. Cell. Proteomics *7*, 2386–2398.

Chen, Y., Chen, W., Cobb, M.H., and Zhao, Y. (2009). PTMap—a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. Proc. Natl. Acad. Sci. USA *106*, 761–766.

Chick, J.M., Kolippakkam, D., Nusinow, D.P., Zhai, B., Rad, R., Huttlin, E.L., and Gygi, S.P. (2015). A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. Nat. Biotechnol. *33*, 743–749.

Choi, H., and Nesvizhskii, A.I. (2008). Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. J. Proteome Res. *7*, 254–265.

Cogliati, S., Calvo, E., Loureiro, M., Guaras, A.M., Nieto-Arellano, R., Garcia-Poyatos, C., Ezkurdia, I., Mercader, N., Vázquez, J., and Enriquez, J.A. (2016). Mechanism of super-assembly of respiratory complexes III and IV. Nature *539*, 579–582.

Craven, L., Tuppen, H.A., Greggains, G.D., Harbottle, S.J., Murphy, J.L., Cree, L.M., Murdoch, A.P., Chinnery, P.F., Taylor, R.W., Lightowlers, R.N., et al. (2010). Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. Nature *465*, 82–85.

Dai, J., Jin, W.H., Sheng, Q.H., Shieh, C.H., Wu, J.R., and Zeng, R. (2007). Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. J. Proteome Res. *6*, 250–262.

Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics *13*, 22–24.

Eng, J.K., Hoopmann, M.R., Jahan, T.A., Egertson, J.D., Noble, W.S., and MacCoss, M.J. (2015). A deeper look into Comet—implementation and features. J. Am. Soc. Mass Spectrom. *26*, 1865–1874.

Ferries, S., Perkins, S., Brownridge, P.J., Campbell, A., Eyers, P.A., Jones, A.R., and Eyers, C.E. (2017). Evaluation of parameters for confident phosphorylation site localization using an orbitrap fusion tribrid mass spectrometer. J. Proteome Res. *16*, 3448–3459.

García-Marqués, F., Trevisan-Herraz, M., Martínez-Martínez, S., Camafeita, E., Jorge, I., Lopez, J.A., Méndez-Barbero, N., Méndez-Ferrer, S., Del Pozo, M.A., Ibáñez, B., et al. (2016). A novel systems-biology algorithm for the analysis of coordinated protein responses using quantitative proteomics. Mol. Cell. Proteomics *15*, 1740–1760.

Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D.L., Dianes, J.A., Del-Toro, N., Rurik, M., Walzer, M.W., Kohlbacher, O., Hermjakob, H., et al. (2016). Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. Nat. Methods *13*, 651–656.

Hämäläinen, R.H., Ahlqvist, K.J., Ellonen, P., Lepisto, M., Logan, A., Otonkoski, T., Murphy, M.P., and Suomalainen, A. (2015).). mtDNA Mutagenesis Disrupts Pluripotent Stem Cell Function by Altering Redox Signaling. Cell Rep. *11*, 1614–1624.

Hämäläinen, R.H., Manninen, T., Koivumaki, H., Kislin, M., Otonkoski, T., and Suomalainen, A. (2013). Tissue- and cell-type-specific manifestations of heteroplasmic mtDNA 3243A>G mutation in human induced pluripotent stem cell-derived disease model. Proc. Natl. Acad. Sci. USA *110*, E3622–E3630.

Hu, C.W., Hsu, C.L., Wang, Y.C., Ishihama, Y., Ku, W.C., Huang, H.C., and Juan, H.F. (2015). Temporal phosphoproteome dynamics induced by an ATP synthase inhibitor citreoviridin. Mol. Cell. Proteomics *14*, 3284–3298.

Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. Cell *143*, 1174–1189.

Jenuth, J.P., Peterson, A.C., Fu, K., and Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat. Genet. *14*, 146–151.

Jin, W.H., Dai, J., Zhou, H., Xia, Q.C., Zou, H.F., and Zeng, R. (2004). Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry. Rapid Commun. Mass Spectrom. *18*, 2169–2176.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. *74*, 5383–5392.

Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277.

Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods *14*, 513–520.

Lee, Y., Stiers, K.M., Kain, B.N., and Beamer, L.J. (2014). Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency. J. Biol. Chem. *289*, 32010–32019.

Leyfer, D., and Weng, Z. (2005). Genome-wide decoding of hierarchical modular structure of transcriptional regulation by cis-element and expression clustering. Bioinformatics *21* (*Suppl 2*), ii197–ii203.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. *17*, 676–682.

Liu, L., and Pilch, P.F. (2016). PTRF/Cavin-1 promotes efficient ribosomal RNA transcription in response to metabolic challenges. eLife *5*, e17508.

López-Carballo, G., Moreno, L., Masiá, S., Pérez, P., and Barettino, D. (2002). Activation of the phosphatidylinositol 3-kinase/Akt signaling pathway by retinoic acid is required for neural differentiation of SH-SY5Y human neuroblastoma cells. J. Biol. Chem. *277*, 25297–25304.

Lundby, A., Andersen, M.N., Steffensen, A.B., Horn, H., Kelstrup, C.D., Francavilla, C., Jensen, L.J., Schmitt, N., Thomsen, M.B., and Olsen, J.V. (2013). In vivo phosphoproteomics analysis reveals the cardiac targets of β-adrenergic receptor signaling. Sci. Signal. *6*, rs11.

Ma, C.W., and Lam, H. (2014). Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. J. Proteome Res. *13*, 2262–2271.

Martínez-Acedo, P., Núñez, E., Gómez, F.J., Moreno, M., Ramos, E., Izquierdo-Álvarez, A., Miró-Casas, E., Mesa, R., Rodriguez, P., Martínez-Ruiz, A., et al. (2012). A novel strategy for global analysis of the dynamic thiol redox proteome. Mol. Cell. Proteomics *11*, 800–813.

Meier, A., and Söding, J. (2015). Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. PLoS Comput. Biol. *11*, e1004343.

Mollapour, M., and Neckers, L. (2012). Post-translational modifications of Hsp90 and their contributions to chaperone regulation. Biochim. Biophys. Acta *1823*, 648–655.

Navarro, P., and Vázquez, J. (2009). A refined method to calculate false discovery rates for peptide identification using decoy databases. J. Proteome Res. *8*, 1792–1796.

Navarro, P., Trevisan-Herraz, M., Bonzon-Kulichenko, E., Núñez, E., Martínez-Acedo, P., Pérez-Hernández, D., Jorge, I., Mesa, R., Calvo, E., Carrascal, M., et al. (2014). General statistical framework for quantitative proteomics by stable isotope labeling. J. Proteome Res. *13*, 1234–1247.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci. Signal. *4*, rs3.

Sharpley, M.S., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C.S., Masubuchi, S., Friend, N., Koike, M., et al. (2012). Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. Cell *151*, 333–343.

Shortreed, M.R., Wenger, C.D., Frey, B.L., Sheynkman, G.M., Scalf, M., Keller, M.P., Attie, A.D., and Smith, L.M. (2015). Global identification of protein post-translational modifications in a single-pass database search. J. Proteome Res. *14*, 4714–4720.

Skinner, O.S., and Kelleher, N.L. (2015). Illuminating the dark matter of shotgun proteomics. Nat. Biotechnol. *33*, 717–718.

van Pée, K.H., and Patallo, E.P. (2006). Flavin-dependent halogenases involved in secondary metabolism in bacteria. Appl. Microbiol. Biotechnol. *70*, 631–641.

Wang, Y.T., Pan, S.H., Tsai, C.F., Kuo, T.C., Hsu, Y.L., Yen, H.Y., Choong, W.K., Wu, H.Y., Liao, Y.C., Hong, T.M., et al. (2017). Phosphoproteomics reveals HMGA1, a CK2 substrate, as a drug-resistant target in non-small cell lung cancer. Sci. Rep. *7*, 44021.

Wiśniewski, J.R., Ostasiewicz, P., and Mann, M. (2011). High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. J. Proteome Res. *10*, 3040–3049.

Wolf, D.P., Mitalipov, N., and Mitalipov, S. (2015). Mitochondrial replacement therapy in reproductive medicine. Trends Mol. Med. *21*, 68–76.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals		
Filter Aided Sample Preparation (FASP) kit	Expedeon	Cat# SKU: 44250
10 plex-TMT	Thermo Fisher Scientific	Cat# 90110
OASIS HLB extraction cartridges	Waters Corp.	Cat# 186000383
high pH reversed-phase peptide fractionation kit	Thermo Fisher Scientific	Cat# PI84868
EASY-nLC 1200 instrument	Thermo Fisher Scientific	Cat# LC120
EASY-Spray NG Ion Source	Thermo Fisher Scientific	Cat# ES082
Q Exactive HF mass spectrometer	Thermo Fisher Scientific	Cat# IQLAAEGAAPFALGMBFZ
Acclaim PepMap100 C18 100 A nanoViper Trap column.	Thermo Fisher Scientific	Cat# 11352013
EASY-Spray column, 50 cm × 75 μm ID, PepMap RSLC C18, 2 μm particles,100 A pore size	Thermo Fisher Scientific	Cat# ES803
Deposited Data		
Raw and analyzed data	This paper	ftp://ftp.peptideatlas.org/ username: PASS01085 password: TC4334fh.
Experimental Models: Organisms/Strains		
Mouse:12-week-old BL/6^{C57}, BL/6^{NZB} and BL/6 $^{C57\ -NZB}$ males.	José Antonio Enríquez Lab, Centro Nacional de Investigaciones Cardiovasculares (CNIC)	BL/6^{C57}, BL/6^{NZB} and BL/6 $^{C57\ -NZB}$
Software and algorithms		
Comet release 2016.01	Eng et al., 2013, 2015	http://comet-ms.sourceforge.net
Comet-PTM	This paper	https://github.com/CNIC-Proteomics/Comet-PTM
SHIFTS (Systematic Hypothesis-free Identification of modifications with controlled FDR based on ultra-Tolerant database Search	This paper	https://github.com/CNIC-Proteomics/SHIFTS
Vseq program	Cogliati et al., 2016	available upon request
Generic Integration Algorithm (GIA)	This paper and García-Marqués et al., 2016	https://github.com/CNIC-Proteomics/PTM-Quant-Stats

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests should be directed to and will be fulfilled by the Lead Contact, Jesús Vázquez (jvaquez@cnic.es).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse Model of Heteroplasmy

All animal procedures conformed to EU Directive 86/609/EEC and Recommendation 2007/526/EC regarding the protection of animals used for experimental and other scientific purposes, enforced in Spanish law under Real Decreto 1201/2005. The mice were fed a standard chow diet (5K67 LabDiet).

Generation of Heteroplasmic Mice

Heteroplasmic mice were generated by electro-fusing cytoplasts from conplastic BL/6^{NZB} zygotes to recipient C57BL/6JOlaHsd (BL/6^{C57}) one cell embryos, cultured overnight and transplanted as two-cell embryos into pseudo pregnant Hsd:ICR (CD-1®) females to complete development to term as previously described (Jenuth et al., 1996). To the best of our knowledge, no consensus rule to name heteroplasmic mouse strains exists. Here, we propose the following designation to name heteroplasmic mouse strains: NUCLEAR GENOME-mtCYTOPLASMIC GENOME #1 + CYTOPLASMIC GENOME #2 [i.e., C57BL/6J-mtC57BL/6+NZB, a strain

with the nuclear genome of C57BL/6J and the cytoplasmic (mitochondrial) genome of C57BL/6J and NZB]. To simplify we are calling it BL/$6^{C57\text{-}NZB}$ along this report. The female heteroplasmic offspring (named BL/$6^{C57\text{-}NZB}$) were mated with C57BL/6JOlaHsd males to prevent nuclear genetic drift in our particular mice lines. Only offspring of the established heteroplasmic mice were used.

Mice Breeding

Heteroplasmic females (BL/$6^{C57\text{-}NZB}$) were outcrossed with males BL/6^{C57}. Only females with an initial level of NZB heteroplasmy above 20% were used for colony maintenance.

The mice used in this work were 12-week-old control (C57BL/6JOlaHsd strain) and heteroplasmic males (containing more than one mtDNA in the same cytoplasm, C57BL/6 background). The effect of heteroplasmy on the PTMs of the proteome of different tissues was studied on liver, heart, and skeletal muscle (gastrocnemius) samples. In the last two tissues, the heteroplasmy was stable, while the liver was selected as a control tissue since it spontaneously selected one of the alternative variants of mtDNA (manuscript in submission For each tissue, biological replicates from different control (N = 3) and heteroplasmic mice (N = 4) were analyzed.

METHOD DETAILS

Benchmarking MS Dataset

To test the performance of the developed algorithms, we used the publicly available HEK293 dataset (Chick et al., 2015), containing 1.121.149 MS/MS spectra in 24 raw files acquired on a Q-Exactive Orbitrap mass spectrometer. For the bench-marking of site localization through CometPTM we used a synthetic phosphopeptide dataset from the Pride database (dataset identifier PXD007058)(Ferries et al., 2017).

Preparation of Protein Extracts

Mice were sacrificed by cervical dislocation and liver, heart, and skeletal muscle tissues were extracted. 20 mg of each tissue were homogenized in lysis buffer (10mM Tris-HCL pH7.4, 1 mM EDTA, 0.32 M sucrose, 2% SDS) freshly supplemented with protease and phosphatase inhibitors (Roche) and 50 mM DTT, using a MagNA Lyser instrument (Roche). The lysate was boiled for 5 min and cell debris were removed by centrifugation.

Protein Digestion, Peptide Labeling, and Fractionation

Proteins were treated with 50 mM iodoacetamide (IAM) and digested with trypsin using the Filter Aided Sample Preparation (FASP) digestion kit (Expedeon) (Wiśniewski et al., 2011) according to manufacturer's instructions. Dried peptides were labeled using 10 plex-TMT reagents according to manufacturer's instructions (Thermo Fisher Scientific), desalted on OASIS HLB extraction cartridges (Waters Corp.) (Leyfer and Weng, 2005), separated into 7 fractions using the high pH reversed-phase peptide fractionation kit (Thermo Fisher Scientific) and dried-down before MS analysis.

LC-MS Analysis

Each fraction of the labeled peptide samples were analyzed using an Easy nano-flow HPLC system (Thermo Fisher Scientific) coupled via a nanoelectrospray ion source (Thermo Fisher Scientific, Bremen, Germany) to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). C18-based reverse phase separation was used with a 2-cm trap column and a 50-cm analytical column (EASY column, Thermo). Peptides were loaded in buffer A (0.1% formic acid (v/v)) and eluted with a 240 min linear gradient of buffer B (80% acetonitrile, 0.1% formic acid (v/v)) at 200 nL/min. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS using a top 15 method. MS spectra were acquired in the Orbitrap analyzer with a mass range of 400–1500 m/z and 60,000 resolution. HCD fragmentation was performed at 27 of normalized collision energy and MS/MS spectra were analyzed at 60,000 resolution in the Orbitrap.

Database Search

Unless indicated otherwise, all searches were performed using Comet release 2016.01 (Eng et al., 2013, 2015) using trypsin digestion with 1 missed cleavages (unless otherwise specified) and fixed Cys carbamidomethylation (57.021464 Da). For heteroplasmic mice data, TMT labeling at N-terminal end and Lys was also considered as a fixed modification (229.162932 Da). Fragment ion tolerance was 0.02 bin, 0 mass offset. Precursor tolerance type and isotope error were set to 1. Precursor charge range was 2-4, maximum precursor charge 5 and maximum fragment charge 3. Only y- and b-ions were used for scoring.

Closed searches (CS) were performed at 5 ppm precursor ion tolerance, using three dynamic modifications: Met oxidation (15.994915), Asn and Gln deamidation (0.984016) and Ser and Thr phosphorylation (79.966331). False discovery rates (FDR) of peptide identifications were calculated using the refined method (Bonzon-Kulichenko et al., 2015; Navarro and Vázquez, 2009); 1% FDR was used as the default criterion for peptide identification.

Open searches (OS) with Comet and Comet-PTM were performed in the same conditions as CS, except that precursor ion tolerance was set to 500 Da.

Comet-PTM

Comet.PTM was developed by modifying the open-source database search engine (Eng et al., 2015). For every sequence candidate Comet-PTM calculates the difference between theoretical and experimental precursor mass (ΔMass), and adds up this mass iteratively to each one of the amino acid masses in the peptide sequence, calculating a Xcorr score in each one of the possible modified forms of the peptide (Figure 1B). The selected candidate is the modified peptide form that produces the highest Xcorr. This design allows Comet-PTM to reach the score that would have been obtained by performing a targeted CS with the same modification in the same position. Note that the scores are not exactly identical, since CS uses the theoretical mass of the modification and Comet-PTM estimates it from the difference between the precursor mass and the theoretical mass of the non-modified peptide, and experimental errors on this estimate may affect fragment matching. This effect is, however, small when low ppm precursor mass accuracies are used (Figures S1A and S1B). Comet-PTM has a user-selectable option of scoring also the non-modified peptide sequence (even when ΔMass is different from zero), to take into account labile modifications (Kong et al., 2017).

Comet-PTM was developed to take full advantage of the multi-thread design of Comet. Comet-PTM used less than 4 hr to perform a 500 Da-wide OS of 16 LC-MS runs, containing an average of 44.390 MS/MS spectra each, using a computer cluster with 16 nodes, where each node is built of 2 x Intel Xeon E5-2695v2 at 2.40 GHz and contained 46 threads/124 gigabyte.

SHIFTS

SHIFTS (Systematic Hypothesis-free Identification of modifications with controlled FDR based on ultra-Tolerant database Search) is a program that identifies peaks in the ΔMass distribution, assigns PSM to peaks and calculates FDR for peptide identification (Figure S6). SHIFTS uses as input the Thermo .raw files and the files obtained from Comet-PTM search.

Mass Recalibration

SHIFTS first recalibrates precursor peptide masses independently in each raw file. This was done by selecting a population of non-modified peptides with a very high score (user selectable; recommended values are those yielding 0.1% global FDR or lower), which are assumed to be true identifications and are used to calculate the systematic mass error (median deviation in m/z scale), which is assumed to be constant in each raw file. From these data, SHIFTS also calculates the standard deviation of the mass error (σ_M) using the median absolute deviation (MAD) method.

Peak Identification

Recalibrated ΔMass values were binned using 0.001 Da bins to construct the ΔMass distribution. The distribution was smoothed using the median of a 7-point sliding window and then peak apexes were detected as downward zero-crossings in the first derivative of the smoothed curve. Peak widths were similarly calculated as the zero-crossing points of the second derivative; in the current version of SHIFTS they are computed only for informative purposes.

Peak Assignation

By default SHIFTS assigns a PSM to the closest ΔMass peak if the mass deviation of the PSM from the peak falls within $3\sigma_M$, so that approximately 99% of PSM in each peak are assigned. This value can be user adjusted. PSM not assigned to peaks were considered as orphan PSM.

FDR Calculation

SHIFTS calculates FDR of identification using a conventional target/decoy strategy using the corrected Xcorr score (cXcorr) (Choi and Nesvizhskii, 2008; Keller et al., 2002). A *global FDR* was calculated for each PSM as the ratio of the number of decoy PSMs to the number of target PSMs having a cXcorr equal or higher. Decoy peptides matched by Comet-PTM were observed to be almost as abundant as target peptides in the negative ΔMass region below the peak corresponding to neutral loss of Gly (Figure 2A), where ΔMass peaks were mostly produced by neutral loss of amino acids. For this reason, the global FDR was only calculated in the ΔMass region above −56 Da (Figure 2A). All the PSM are required to have FDR lower than the global FDR, without exception.

In addition, local FDR filters are also applied. Some ΔMass peaks were observed to contain an unusually high number of decoy PSM; to avoid matching false positive target PSM in these peaks, SHIFTS also calculates a *peak FDR* counting up the number of decoys and target PSM assigned to each peak, and these PSM are required to pass the peak FDR filter in addition to the global FDR filter (Figure 2C). Note that peak FDRs are often very low suggesting that the majority of PSM in these peaks are true, even when they have a low cXcorr. This happens because the probability of finding a decoy PSM in a peak by chance alone is extremely low. SHIFTS avoids matching these low scoring target PSM by applying the global FDR filter.

To apply a local filter to PSM which are not assigned to ΔMass peaks, e.g., to orphan PSM, SHIFTS models the periodic mass distribution of decoy PSM into ~1 Da-bins centered at the regions where ΔMass values concentrate, and calculates a *local FDR* by counting up decoy and target PSM in each one of these regions (Figure 2B). The local FDR filter is applied to orphan PSM in addition to the global FDR filter. Default values for peptide identification were 1% for peak and local FDR; 5% for global FDR. Among the peptides significantly changed by heteroplasmy and suspected to have relevant biological activity, including all the peptides in Table S3, only those that passed analysis with Vseq program (Cogliati et al., 2016) were considered as trustable identifications.

Isotopic Correction

SHIFTS also performs a simple isotopic correction to minimize misassignations of the correct monoisotopic peak of the precursor. When two PSM having the same sequence are encountered having a ΔMass difference within 1 ppm of the mass difference expected for either one or two ^{13}C or one ^{34}S, the ΔMass of the heaviest precursor is substituted by that of the lightest one.

Annotation of Modifications

A Python in-house script was used for semisupervised annotation of the nature of peptide modifications. The script searched ΔMass values against Unimod database, taking into account the amino acid modified according to Comet-PTM output and also the preceding and consecutive residues, comparing them with the list of amino acids that could be subjected to the modification according to Unimod. If no amino acid was matched, the modification was considered as unassigned. Residues containing modifications that were fixed in the database search were considered with and without the fixed modification; when not indicated, the amino acid contains the fixed modification (e.g., C_oxidation means oxidation of carbamidomethylated Cys and K_oxidation, oxidation of TMT-labeled Lys). ΔMass values that could not be matched were tentatively tested assuming one ^{13}C misassignment of the monoisotopic mass of the precursor and also as combinations of two modifications from a list of the most abundant modifications found in the corresponding proteome. Unexplained ΔMass values were termed as unknown. MS/MS fragmentation spectra from the most abundant modifications that changed their abundance in heteroplasmic mice and all the peptides in Table S3 were revised using Vseq program (Cogliati et al., 2016).

QUANTIFICATION AND STATISTICAL ANALYSIS

Peptide Quantification and Statistical Analysis

The quantitative information from TMT reporter intensities was integrated from the spectrum level to the peptide level and then to the protein level on the basis of the WSPP model (Martínez-Acedo et al., 2012; Navarro et al., 2014) using the Generic Integration Algorithm (GIA) (García-Marqués et al., 2016) . The algorithm was modified to include into the statistical model the quantitative values of modified peptides as part of the automated workflow (Figure 3A). Briefly, for each sample i the values $x_{qps} = \log_2 S_i/C$ were calculated, where S_i is the intensity of the TMT reporter corresponding to sample i in the MS/MS spectrum s coming from peptide p and protein q, and C is the average intensity of all the TMT reporters from the control samples, which is used as a common reference. The $\log_2$-ratio of each peptide (x_{qp}) was calculated as the weighted average of its spectra, the protein values (x_q) were the weighted average of its peptides, and the grand mean $(\overline{x})$ was calculated as the weighted average of all the protein values (Navarro et al., 2014) . The statistical weights of spectra, peptides, and proteins (w_{qps}, w_{qp} and w_q, respectively) and the variances at each one of the three levels ($\sigma^2_{\mathbf{S}}$, $\sigma^2_{\mathbf{P}}$, and $\sigma^2_{\mathbf{Q}}$, respectively), were calculated as described (Navarro et al., 2014) .

The spectrum, peptide, and protein variances and the protein values were first determined including only non-modified peptides (Figure 3B). In a second step, the modified peptides were included in the analysis, which was performed using the variances and protein values calculated previously. For each modified peptide, the standardized variable (z_{pq}) was calculated as

$$z_{pq} = (x_{qp} - x_q)\sqrt{w_{pq}}\sqrt{\frac{n_p}{n_p - 1}},\ n_p > 1$$

where n_p is the number of non-modified peptides with which the corresponding protein was quantified. z_{pq} expresses the deviation between the peptide $\log_2$-ratio and the corresponding protein value in units of standard deviation. In absence of changes, the distributions of z_{pq} followed very closely the normal distribution $N(0,1)$ (Figure 3C), validating the accuracy of the model.

Significant abundance changes of modified peptides were detected by Student's t test comparing the z_{pq} values of samples from heteroplasmic mice (N = 4) with those of control mice (N = 3) (Table S2).

DATA AND SOFTWARE AVAILABILITY

Code Access

The software needed to execute the whole Comet-PTM pipeline can be downloaded from: https://github.com/CNIC-Proteomics/Comet-PTM.

The software for Comet-PTM FDR control can be downloaded from: https://github.com/CNIC-Proteomics/SHIFTS.

The software for statistical analysis of quantitative data can be downloaded from https://github.com/CNIC-Proteomics/PTM-Quant-Stats. A readme.txt file is provided with basic instructions to install and execute the package. Vseq is available upon request.

The dataset of the raw files, protein databases, search parameters and results reported in this paper is available in the PeptideAtlas repository at ftp://ftp.peptideatlas.org/ (username: PASS01085, password: TC4334fh).

3. Publication related to PhD work

3.2. Proteomic footprint of myocardial ischemia/reperfusion injury: *longitudinal study of the at-risk and remote regions in the pig model*

SCIENTIFIC REPORTS

OPEN

Proteomic footprint of myocardial ischemia/reperfusion injury: *Longitudinal study of the at-risk and remote regions in the pig model*

Received: 25 April 2017
Accepted: 25 August 2017
Published online: 27 September 2017

Aleksandra Binek[1], Rodrigo Fernández-Jiménez[1,2,3], Inmaculada Jorge[1,3], Emilio Camafeita[1,3], Juan Antonio López[1,3], Navratan Bagwan[1], Carlos Galán-Arriola[1,3], Andres Pun[1,3], Jaume Agüero[1,3], Valentin Fuster[1,2], Borja Ibanez[1,3,4] & Jesús Vázquez[1,3]

Reperfusion alters post-myocardial infarction (MI) healing; however, very few systematic studies report the early molecular changes following ischemia/reperfusion (I/R). Alterations in the remote myocardium have also been neglected, disregarding its contribution to post-MI heart failure (HF) development. This study characterizes protein dynamics and contractile abnormalities in the ischemic and remote myocardium during one week after MI. Closed-chest 40 min I/R was performed in 20 pigs sacrificed at 120 min, 24 hours, 4days, and 7days after reperfusion (n = 5 per group). Myocardial contractility was followed up by cardiac magnetic resonance (CMR) and tissue samples were analyzed by multiplexed quantitative proteomics. At early reperfusion (120 min), the ischemic area showed a coordinated upregulation of inflammatory processes, whereas interstitial proteins, angiogenesis and cardio-renal signaling processes increased at later reperfusion (day 4 and 7). Remote myocardium showed decreased contractility at 120 min- and 24 h-CMR accompanied by transient alterations in contractile and mitochondrial proteins. Subsequent recovery of regional contractility was associated with edema formation on CMR and increases in inflammation and wound healing proteins on post-MI day 7. Our results establish for the first time the altered protein signatures in the ischemic and remote myocardium early after I/R and might have implications for new therapeutic targets to improve early post-MI remodeling.

Widespread implementation of reperfusion strategies has dramatically reduced mortality associated with myocardial infarction (MI). One consequence of increased survival is an increased prevalence of chronic heart failure (HF), because patients surviving the acute episode then live with a significantly damaged heart[1,2]. Treatments for post-MI HF have primarily focused on symptom management after the occurrence of irreversible remodeling and functional impairment of the left ventricle (LV)[3]. A better understanding of the molecular mechanisms driving post-ischemia/reperfusion (I/R) cardiac dysfunction at earlier stages might enable the development of therapeutic interventions to prevent the onset of HF[4,5].

Much of the prevailing knowledge about post-infarcted tissue and the molecular modifications involved comes from studies performed in non-reperfused conditions[6–8]. Surprisingly, although reperfusion is known to alter post-MI myocardial healing[9,10], to date there have been no systematic studies on the molecular changes occurring in the post-reperfused myocardium at early stages after I/R. Moreover, previous proteomic characterizations of post-I/R myocardial tissue were limited to mitochondrial proteins[11] or extracellular matrix proteins[12] or were performed at single time points, mostly at late remodeling stages (weeks after MI)[12–15]. There has also been a notable lack of attention paid to the remote myocardium, which has been used in most studies as a control tissue for the analysis of changes in ischemic areas, thus carrying the implicit assumption that no important changes

[1]Fundación Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain. [2]The Zena and Michael A. Wiener CVI, Icahn School of Medicine at Mount Sinai, New York, USA. [3]CIBER de Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain. [4]IIS-Fundación Jiménez Díaz Hospital, Madrid, Spain. Aleksandra Binek and Rodrigo Fernández-Jiménez contributed equally to this work. Correspondence and requests for materials should be addressed to B.I. (email: bibanez@cnic.es) or J.V. (email: jvazquez@cnic.es)

take place in this area. Unsurprisingly therefore, the molecular alterations taking place in the remote myocardium at early post-MI stages are poorly understood. This lack of knowledge is particularly important because the remote myocardium plays an important role in the development of maladaptive LV remodeling and subsequent HF after MI[16,17].

In a pig model of myocardial I/R, serial cardiac magnetic resonance (CMR) imaging shows that tissue composition changes dynamically during the first week after infarction[18]. After a transient hyperacute, reperfusion-related edematous reaction, there is a longer, healing-related deferred phase of edema[10]. The precise molecular and cellular events taking place during these highly dynamic early post-I/R phases remain largely unknown.

In this study, we characterize the dynamic changes that take place in both the ischemic and remote myocardium during the first week after infarction in a clinically relevant animal model. This was undertaken by high-throughput multiplexed quantitative proteomics of pig myocardium at different times during the first week after I/R, together with serial *in-vivo* tissue characterization by CMR at each time point. In a first screening approach, we provide a detailed pattern of the time-course behavior of more than 5.000 proteins in both the ischemic and remote areas. This was followed by a further validation step in which we focused on most relevant protein changes. Our results reveal a highly coordinated, multimodal pattern of functional protein alterations in both myocardium regions in the early stages post infarction. Most notably, the remote myocardium undergoes transitory alterations in the contractile machinery that result in a stunned myocardium within the first 24 h. Addressing the need for unbiased "omics" approaches, our data constitute a rich resource of information about the precise molecular signatures and pathways that characterize early remodeling stages, with potential utility for molecular researchers looking for novel drug targets for cardioprotection and HF prevention[5].

Results

Time profile of regional myocardial contractility during early post-I/R remodeling. Five groups of animals (n = 5 per group) were used in this study (Fig. 1a): (1) control pigs (not undergoing I/R), (2) I/R pigs sacrificed 120 min after reperfusion, (3) I/R pigs sacrificed 1 day after reperfusion, (4) I/R pig sacrificed 4 days after reperfusion, and (5) I/R pigs sacrificed 7 days after reperfusion. Table 1 summarizes functional cardiac CMR data obtained during early remodeling after I/R. As expected, regional contractility in the post-ischemic myocardium, as measured by systolic wall thickening (SWT), was impaired at all time-points evaluated. The remote myocardium also showed a significant transitory alteration in regional myocardial function, with SWT significantly reduced at 120 min and 24 h, and then recovering to baseline values at day 7 post-reperfusion. The SWT time course in the ischemic and remote myocardium is shown in Fig. 2, together with a representative bullseye display of SWT of one pig at all time-points evaluated.

Proteome changes in the ischemic myocardium during early stage remodeling after I/R. In a first screening approach, we subjected pooled heart extracts from five animals per group to high-throughput multiplexed quantitative proteomics in three technical replicates. This allowed very precise relative quantification of myocardial proteins with very high proteome coverage. We were able to quantify more than 5000 proteins from the ischemic and remote myocardium. Functional annotation was followed by a novel systems biology analysis developed by our group[19]. This analysis detected statistically significant biological changes (FDR < 0.05), specifically driven by the coordinated action of protein subsets that accurately depict phenotypic alterations in myocardial tissue.

Early after reperfusion (120 min), the ischemic myocardium showed a coordinated increase in the expression of proteins involved in acute-phase response signaling, response to wounding, wound healing, blood cell adhesion, and production of nitric oxide and reactive oxygen species (Fig. 3a), and this was accompanied by decreases in proteins involved in glycolysis, alcohol biosynthesis, and kinase binding (Fig. 3b). These processes reflect an early activation of immunological/inflammatory responses in the post-reperfused myocardium. In addition, there was also a coordinated decrease in cell junction proteins (Fig. 3b), which is consistent with the initial wave of edema formation described in this model[18].

At a later stage after reperfusion (day 4 to day 7), the ischemic myocardium showed a coordinated increase in protein biosynthesis processes, manifested by increased abundance of proteins related to gene expression, ribosomes, and vesicle-mediated protein transport (Fig. 3c). There was also significant increase of collagen and extracellular matrix proteins, peaking at day 7 post-reperfusion, and of angiogenesis-associated proteins, starting at day 4 and peaking at day 7 (Fig. 3c). The category of proteolysis also increased; although this change was less statistically significant, it was consistently increased in the four time points and included a considerable number of proteins (183) (Fig. 3c). These changes correlated with significant fibrosis[10] and are consistent with a late and gradual activation of healing processes. Conversely, between day 4 and day 7 post I/R the ischemic myocardium showed a generalized decrease of proteins involved in the mitochondrial electron transport chain (ETC) (mitochondrial complexes I-V) (Supplementary Figure S1), fatty acid biosynthesis, amino acid metabolism, fatty acid beta-oxidation, tricarboxylic acid (TCA) cycle, and mitochondrial ribosomes (Fig. 3d), suggesting a diminished mitochondrial content. A coordinated decrease in muscle contraction was also detected (Fig. 3d). These changes seem to reflect the removal of necrotic myocytes (rich in mitochondria) and their replacement by fibroblasts (less mitochondrial density) during healing. Interestingly, this stage was also characterized by a coordinated increase in proteins implicated in acute renal failure and in renal toxicity biomarkers (Fig. 3c), indicating concomitant gradual activation of cardio-renal signaling processes. The coordinated nature of the protein responses detected in this analysis is illustrated in Fig. 3e. A representational diagram that relates the changes in water content with the protein clusters in the ischemic myocardium is shown in Fig. 3f.

In the next step of this study, we used alternative approaches to validate the main phenotypic changes. In this step we wanted not only to validate the previous results, but also to determine whether the protein alterations

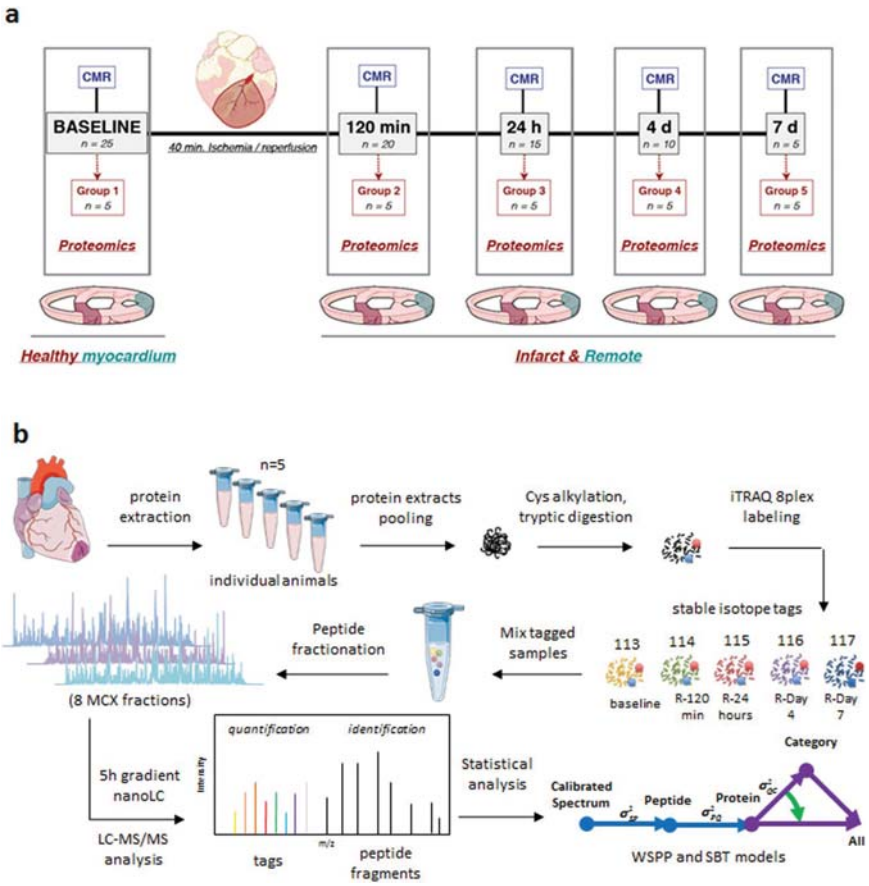


Figure 1. Study design. (**a**) The study population comprised 5 groups of pigs (n = 5/group). Groups 1 to 5 were used for the histopathological characterization of myocardial tissue changes during the first week after ischemia/reperfusion (I/R). CMR (cardiac magnetic resonance) scans included cine steady-state free-precession (SSFP) sequences to provide high quality anatomical references and functional information. CMR scans were performed at all follow-up stages until sacrifice, so that animals sacrificed at day 7 underwent baseline, 120 min, 24 h, day 4, and day 7 CMR. CMR and myocardial water content data from animals in groups 1 to 5 have been reported previously[18]. (**b**) For proteomic analysis, remote and ischemic tissue samples from these animals were processed for protein extraction, tryptic digestion, multiplexed stable isotope labeling, and fractionation followed by nano-liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) and systems biology analysis. Graphical elements in panel b were adapted from the Servier Medical Art Powerpoint image bank. Servier Medical Art by Servier (http://www.servier.com/Powerpoint-image-bank) is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

are consistently reproduced in different individuals. Western blotting analysis of a set of representative proteins from complex I (NDUFB8), complex II (SDHB), complex III (UQCRC2, UQCRFS1), complex IV (COXIVI1) and complex V (ATP5A, ATP5B) from ETC confirmed that the decrease in mitochondrial proteins at 7 days after reperfusion was a generalized event suffered by practically all animals analyzed (n = 5) (Fig. 4a). Similar results were obtained when representative contractile proteins from thick filaments (myosin) and from thin filaments (TNNT2, TPM1) were analyzed by western blot (Fig. 4b). Furthermore, label-free mass spectrometry analysis in individual animals (n = 5) also confirmed that the absolute amounts of proteins, expressed as iBAQ intensities[20], belonging to mitochondria or implicated in contractile processes were decreased (Figs 4c,d, S2 and S13a,b), while acute phase response proteins were clearly increased at 7 days after reperfusion (Figs 4d, S2 and S13c). The total amount of collagens or extracellular matrix proteins were also confirmed to increase (Fig. 4d, S2 and S13d,e). In addition, both the total amount of proteolytic and gene expression proteins were consistently found increased in the animals after 7 days of reperfusion (Figs 4d, S2 and S13f,g). These results suggested that the increase in collagen and extracellular matrix proteins was the result of a net balance between increased mRNA expression and increased proteolytic activity, as proposed by other authors[21].

		Baseline	R-120 min	R-24 hours	R-Day 4	R-Day 7
Group 1 (Control)	**I**	75.7 (16.7)				
	R	61.8 (20.1)				
Group 2 (I/R-120min)	**I**	38.4 (10.2)	−21.4 (15.1)			
	R	18.3 (7.0)	−12.3 (9.3)			
Group 3 (I/R-24hour)	**I**	47.5 (11.6)	−15.8 (12.4)	−4.3 (9.7)		
	R	30.6 (14.0)	6.7 (16.1)	9.7 (13.9)		
Group 4 (I/R-4days)	**I**	45.4 (17.7)	−14.6 (9.7)	−3.7 (7.9)	12.9 (16.9)	
	R	35.3 (16.3)	−5.5 (9.4)	7.9 (21.5)	39.7 (21.7)	
Group 5 (I/R-7days)	**I**	80.6 (16.2)	−10.2 (18.6)	3.6 (15.3)	21.6 (7.8)	17.6 (15.5)
	R	65.6 (14.9)	14.1 (11.3)	18.6 (21.3)	40.0 (12.7)	35.4 (19.7)
Pooled	**I**	**57.5 (22.1)**	**−15.5 (13.8)***	**−1.5 (11.2)***	**17.2 (13.2)***	**17.6 (15.5)***
	R	**42.3 (23.3)**	**0.7 (15.2)***	**12.0 (18.4)***	**39.8 (16.7)**	**35.4 (19.7)**

Table 1. Left ventricular regional systolic wall thickening analysis in the ischemic and remote pig myocardium during the first week after I/R. Left ventricular regional systolic wall thickening in the ischemic and remote myocardium, reported as mean % (standard deviation). A priori pairwise comparisons between baseline and pooled data for each time point were performed using generalized linear mixed models. Multiple comparison adjustment of p-value was performed according to the sequential Holm-Bonferroni procedure. *Statistically significant differences ($p < 0.05$) between baseline and corresponding pooled time-point data. The different myocardial states (ischemic vs. remote) were initially defined by the localization relative to late gadolinium enhanced defined infarction[41,42]. In the case of baseline imaging, i.e. before ischemia/reperfusion, such a definition was retrospectively performed based on follow-up imaging; while matched areas were selected in the case of controls. I: Ischemic; R: Remote

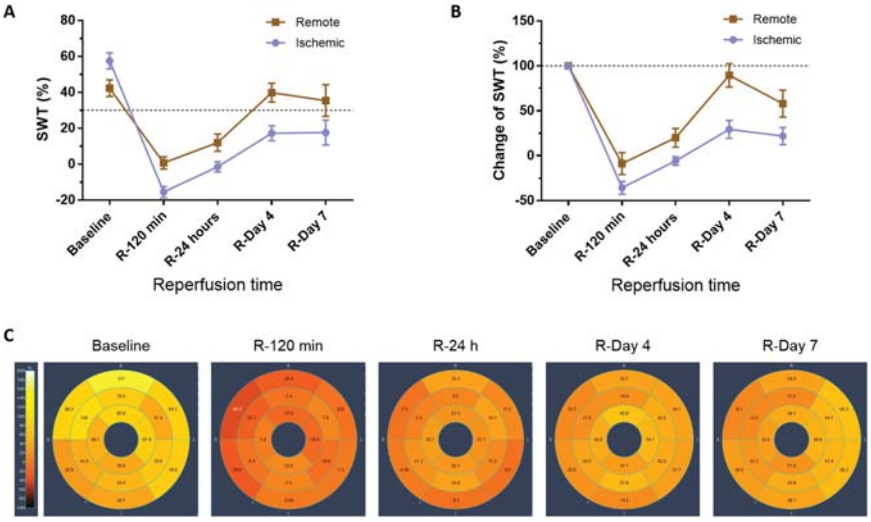


Figure 2. Functional cardiac magnetic resonance during early post-I/R remodeling. Time course of regional contractility in the ischemic and remote myocardium. Systolic wall thickening (SWT) in the ischemic and remote myocardium: (**A**) Absolute values. (**B**) Percentage SWT change relative to baseline. Symbols and bars denote mean and standard error of the mean. Dashed lines indicate reference values from healthy pigs, before induction of myocardial infarction. Myocardial segments were considered dysfunctional if SWT is 30% or less. (**C**) Representative bullseye display of SWT in one pig at baseline and over the first week post I/R after 40-minute mid left anterior descending coronary artery occlusion. Transient contractile dysfunction was observed in the remote myocardium at early post-I/R phases.

Proteome changes in the remote myocardium during early stage remodeling after I/R. At early post-reperfusion stages (120 min to 24 h) the remote myocardium showed a transient coordinated upregulation of proteins related to ATPase activity, muscle thin filament (mainly troponins, tropomyosins, and tropomodulin), proteolysis, fibrillar collagen deposition, and mitochondrial ribosomes (Fig. 5a). Conversely, there was a transient downregulation of muscle thick filament proteins, accessory contractile proteins, and proteins related to several mitochondrial processes (including carriers, tricarboxylic acid cycle, fatty acid beta-oxidation, and ETC complexes, mainly affecting complexes III and IV) (Figs 5b, S3 and S4). With the exception of fibrillar collagen,

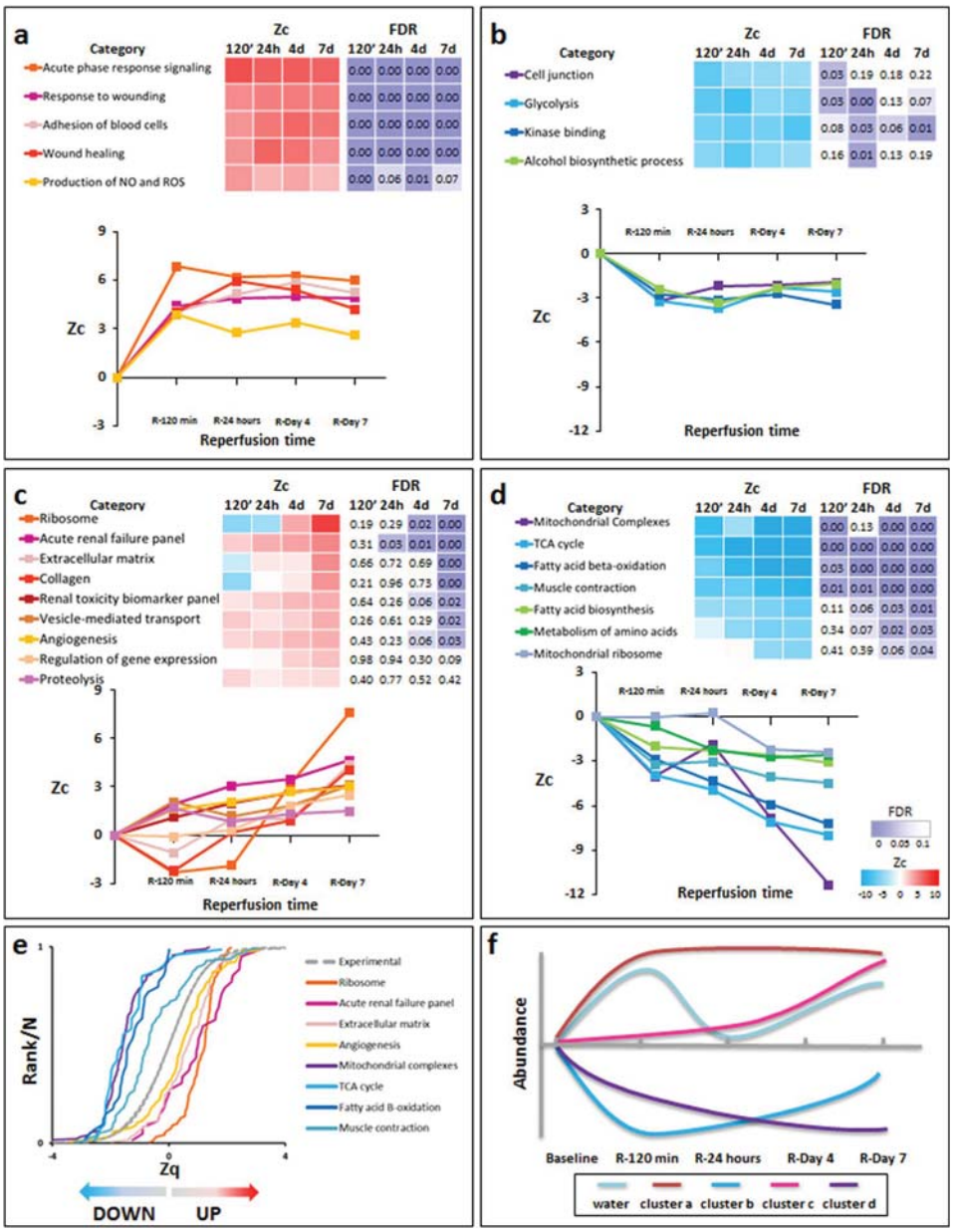


Figure 3. Quantitative proteomics time course analysis of early reperfusion after acute infarction in the ischemic myocardium. Quantitative proteomics screening approach results were analyzed using the Systems Biology Triangle (SBT) model to detect coordinated protein changes in functional categories over time. Functional categories significantly altered (False Discovery Rate, FDR < 0.05) at least one time point were classified into four clusters (**a**–**d**) Cluster a, increasing after 120 min with sustained upregulation for 7 days after ischemia onset; Cluster b, decreasing after 120 min and recovering to baseline values at day 7, except for glycolysis and kinase-binding proteins. Cluster c, increasing at days 4 and 7, with maximum change at day 7; and Cluster d, decreasing over time, with a tendency to reach minimum values at day 7. Protein values (Zq) and functional category values (Zc) are reported as $\log_2$ fold changes with respect to baseline, in units of standard deviation. The complete set of proteins changing in each category is listed in Supplementary Tables S1–S4. Panel (e) displays the cumulative distributions of Zq from proteins belonging to a set of representative categories from clusters **c** and **d**, showing the high coordination of protein responses. (**f**) Schematic chart of changes in water content and protein clusters. The proteomics results presented here are representative of three technical replicate experiments using five animal sample pools.

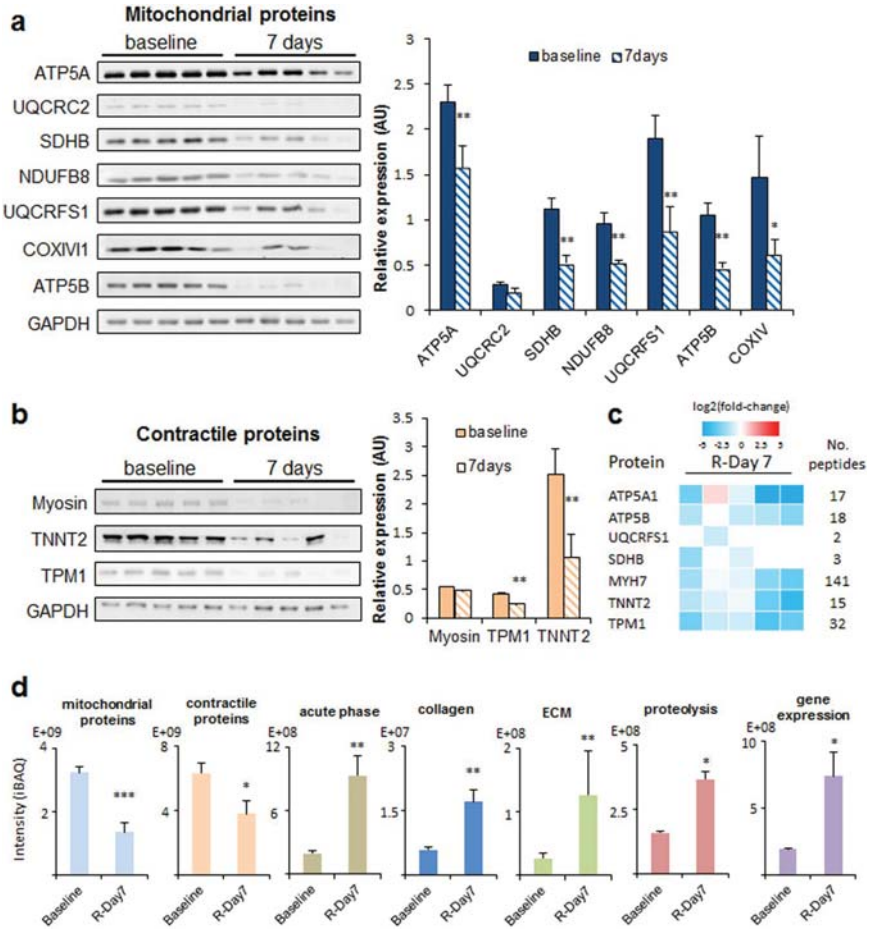


Figure 4. Validation of the most representatitve protein changes detected in ischemic myocardium after acute infarction. (**a**) Western blot analysis and corresponding quantitative densitometry of ATP5A, UQCRC2, SDHB, NDUFB8, UQCRFS1, COXIV and ATP5B confirming the decrease in mitochondrial protein content 7 days after reperfusion with respect to baseline. (**b**) Western blot analysis and quantitative densitometry of cardiac Myosin heavy chain, TNNT2 and TPM1 confirming the decrease in sarcomeric contractile proteins 7 days after reperfusion. Each lane of western blot corresponds to a protein extract obtained from one animal. Protein values are normalized relative to GAPDH. In the bar plots, results are shown as the mean ± s.e.m. of five determinations. *Indicates p value < 0.05; **indicates p value < 0.01 versus baseline by Mann-Whitney test. Full-length blots from panels a and b are presented in Supplementary Figure S8 as well as their raw integrated densities before and after normalization can be consulted in Supplementary Figures S9 and S10. iTRAQ quantifications of the proteins from panels a and b are shown in the Supplementary Figure S11a for comparison. (**c**) Label free individual protein quantifications of ATP5A, ATP5B, UQCRFS1, SDHB, MYH7, TNNT2 and TPM1 confirming the decrease in mitochondrial and contractile protein content. Protein quantifications are represented as log2(fold-change) of iBAQ intensities of individual animals from R-Day 7 compared with pooled average of iBAQ intensities from all Baseline group animals. (**d**) Label free quantitative proteomics confirming the decrease in mitochondrial and contractile proteins and the increase in acute-phase related proteins, collagens, extracellular matrix proteins (ECM), proteolysis related proteins and gene expression related protein contents, 7 days after reperfusion with respect to the baseline. Protein abundances were expressed as the iBAQ values. Results are shown as the mean ± s.e.m of five determinations performed in individual animals, *indicates FDR < 0.05; **indicates FDR < 0.01, ***indicates FDR < 0.001 versus baseline by Mann-Whitney test after Benjamini–Hochberg correction. Label free individual protein quantifications included in the mitochondrial, contractile, acute-phase, collagens, ECM, proteolysis and gene expression protein groups are shown in Supplementary Figure S13.

values for these coordinated processes returned to normal by day 7 after reperfusion (Figs 5a,b and S3). The imbalance in the proportion of sarcomere components strongly suggests uncoupling of the finely-tuned contractile machinery in the remote myocardium early after myocardial infarction. Notably, unlike the pattern seen

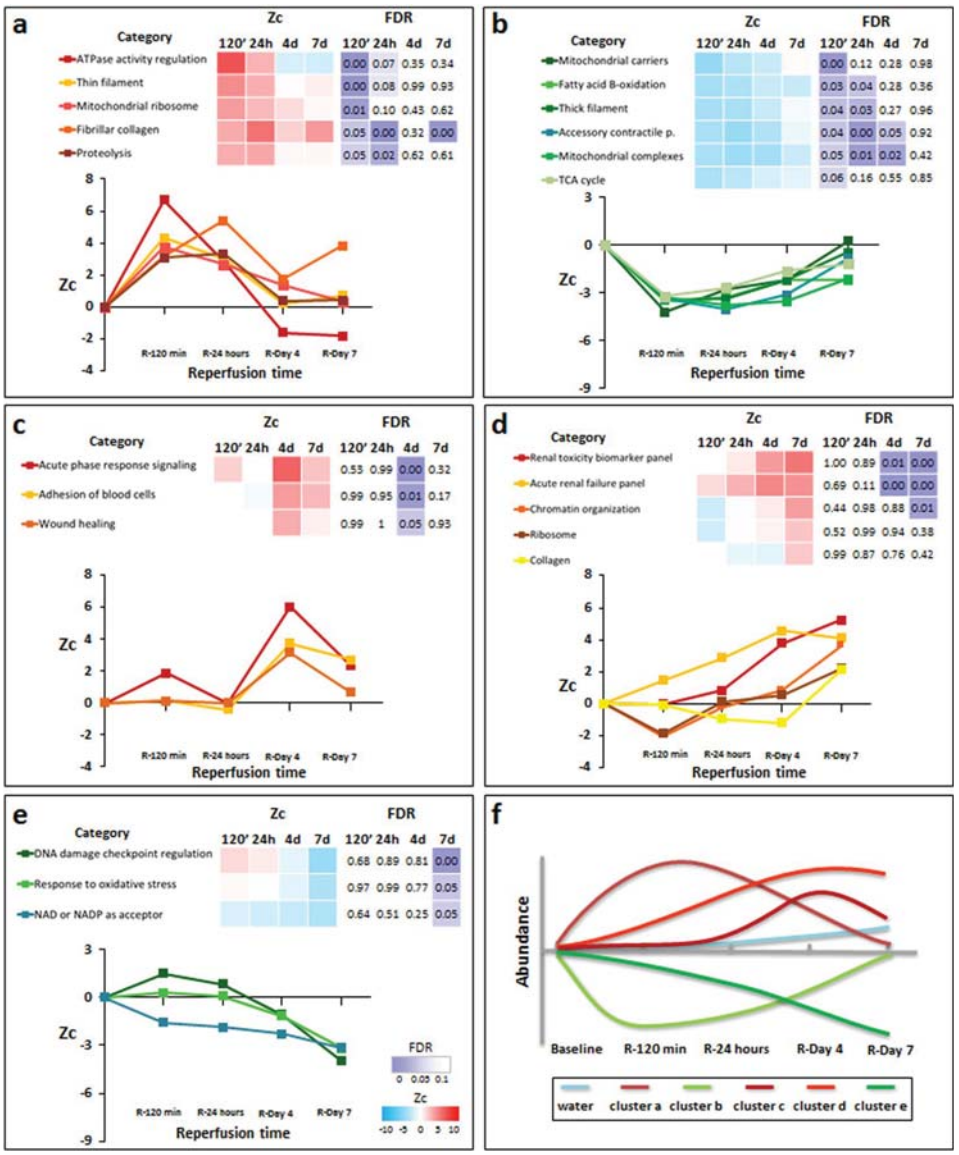


Figure 5. Quantitative proteomics time course analysis of early reperfusion after acute infarction in the remote myocardium. Quantitative proteomics screening approach results are presented as in Fig. 3. Significantly altered functional categories were classified into five clusters (**a**–**e**): Cluster a, showing a transient increase at 120 min or 24 h; Cluster b, decreasing at 120 min and 24 h and reverting to almost normal values at day 7; Cluster c, reaching maximum upregulation at day 4, Cluster d, increasing gradually over time (this cluster included ribosome and collagen categories for comparison of results with the infarcted area); and Cluster e, decreasing in the late phase of reperfusion, reaching minimal values at day 7. Panel (f) shows a schematic chart of changes in water content and protein clusters. The complete set of proteins changing in each category is listed in Supplementary Tables S5–S9. The proteomics results presented here are representative of three technical replicate experiments using five animal sample pools.

in the infarcted area (Figs 3d, 4a,c and S1), mitochondrial proteins in the remote myocardium did not all follow the same trend (Figs 5a,b and S4). This finding suggests that mitochondrial protein alterations do not reflect a decrease in mitochondrial number, but rather a progressive dysfunctional imbalance. Furthermore, this transient uncoupling of the mitochondrial and contractile protein machinery correlated with the temporary contractile dysfunction (SWT) observed by CMR during the first 24 h of reperfusion (Fig. 2). These findings clearly indicate that the molecular alterations triggered by I/R have a transient impact on remote myocardium function early during reperfusion.

Functional analysis in the remote myocardium in late reperfusion phases (4 and 7 days post I/R) revealed an increase in proteins related to acute phase response signaling, wound healing, and blood cell adhesion (Fig. 5c and d). These functional alterations correlated with the linear edema in this area, revealed by progressively increasing CMR T2 relaxation times during the first week after reperfusion[18,22]. Furthermore, there were also increases in proteins related to chromatin organization and cardio-renal signaling processes (acute renal failure and renal toxicity biomarkers), reaching peak values on days 4 and 7 post-I/R (Fig. 5d). Collagen and ribosome-related proteins showed a similar trend, although with lower statistical significance. Downregulated proteins included NAD and NADP acceptor proteins and proteins implicated in oxidative stress and DNA damage (Fig. 5e). Finally, the biological processes grouped in Fig. 5c and d show intriguing time offset/timeline discrepancies compared with the observations in the ischemic myocardium. A schematic diagram of changes in water content and protein categories clusters which occur in the remote myocardium is represented in Fig. 5f.

The main phenotypic changes in remote myocardium were also validated by alternative analysis. Western blot results obtained in at least four animals confirmed the generalized transitory increase at 120 min after reperfusion of a contractile protein belonging to thin filaments (TPM1) (Fig. 6a). Label-free mass spectrometry of individual animal samples (at least n = 4 per each condition) also confirmed that the absolute amount of proteins belonging to thick filaments was decreased at 120 min, while those belonging to thin filaments were increased (Figs 6b, S5 and S14). A more accurate quantification of specific groups of proteins was also performed in individual animal samples (at least n = 4 per each condition) by targeted proteomics, performing two technical replicates per sample. These analysis confirmed the transient decrease at early times (120 min to 1 day) followed by subsequent recovery (7 days after reperfusion) of proteins belonging to mitochondrial carriers (ANT2), TCA cycle (ACO2, MDH), complex V (ATP5A1, ATP5O), complex III (UQCRC1), complex IV (COX6B), thick filaments (MYH7) or the accessory contractile proteins category (MYOM1). Similarly, we could also confirm the transient increase at early times of proteins implicated in ATPase activity regulation (MYL3) or belonging to thin filaments (TPM1). The protein used as internal control (GAPDH) was not significantly altered (Figs 6c and S6). Finally, as stated in the previous paragraphs, in our exploratory iTRAQ experiments of the remote myocardium tissue, we detected a coordinated initial disengagement (at 120 min) of the values of mitochondrial and sarcomeric proteins followed by the latter coordinated restoration to normal (at day 7), which is shown in Fig. 7a-left and Fig. 7a-right, respectively.

Discussion

In the present study, advanced proteomic analysis in a pig model of I/R provides quantitative time-course information on thousands of proteins and unravels an array of coordinated early changes in the heart proteome after MI. We have assessed not only the particular altered protein signatures that characterize the post-reperfused ischemic and remote myocardium, but also their relation to functional changes detected by CMR. This is the first study reporting a comprehensive proteomics analysis of molecular changes during early reperfusion in the post-MI remote myocardium.

Early reperfusion is characterized by an abrupt edematous reaction[10,18] and a rapid and profound acute inflammatory response[23,24]. Upon reperfusion, intracellular, mitochondrial and sarcolemma Ca^{2+} concentrations increase abruptly, inducing cardiomyocyte hypercontraction and opening of the mitochondrial permeability transition pore (MPTP), the hallmark of reperfusion-induced cardiomyocyte death[25]. MPTP opening collapses the mitochondrial membrane potential and disconnects the electron transport chain, resulting in ATP depletion and subsequent cell death[26]. Our deep proteomics analysis of ischemic myocardium fully supports and extends these observations by establishing the time-course of hundreds of protein changes and by demonstrating a coordinated activation of proteins implicated in inflammation, wound healing, and production of reactive oxygen species as soon as 120 min after reperfusion. Furthermore, at these early time-points, we identified downregulation of cell junction and glycolytic metabolism proteins, consistent with the reperfusion-induced interstitial edema[10] and the disruption of the normal mitochondria function[11,27]. Our study identifies a rich repository of early-response proteins directly related to reperfusion injury that could contribute to the development of therapies and diagnostic tools[5].

At later reperfusion phases (4 and 7 days post reperfusion), coordinated protein changes in the ischemic myocardium indicated activation of protein biosynthesis, synthesis of collagen and interstitial proteins, and formation of new vasculature, together with decreases in mitochondrial and muscle proteins. These processes reflect a phenotypic change in the infarcted myocardial tissue caused by the replacement of necrotic myocytes (rich in mitochondria and with high metabolic activity) by collagen, extracellular matrix proteins, and fibroblasts (lower mitochondrial content and metabolic activity). Notably, in the ischemic myocardium we also identified a progressive increase in proteins related to cardio-renal signaling, suggesting active cross-talk between the acutely injured heart and the kidney.

The remote myocardium plays an important role in ventricular function after AMI, and acute post-AMI HF can be exacerbated by transitory stunning of the non-ischemic region[28]. Similarly, the long-term inflammatory response in the remote region can lead to progressive adverse remodeling, chronic ventricular dilatation, and HF[17]. Therefore, understanding the molecular changes occurring in the remote region is of the utmost importance for the development of therapeutic strategies targeting this region. However, the early post-infarct behavior of the remote tissue, particularly at the molecular level, has remained unexplored until now. In this study, we contribute to fill this gap of knowledge by providing a first insight into the transiently stunned myocardium and a highly detailed characterization of phenotypic events taking place in the remote area during the first week after I/R.

We demonstrate that I/R significantly impacts the remote myocardium, inducing molecular changes that correlate precisely with CMR-evaluated alterations in contractile function (transient stunned myocardium) and tissue composition (late progressive edema). In the period from 120 min to 24 h after reperfusion, transient

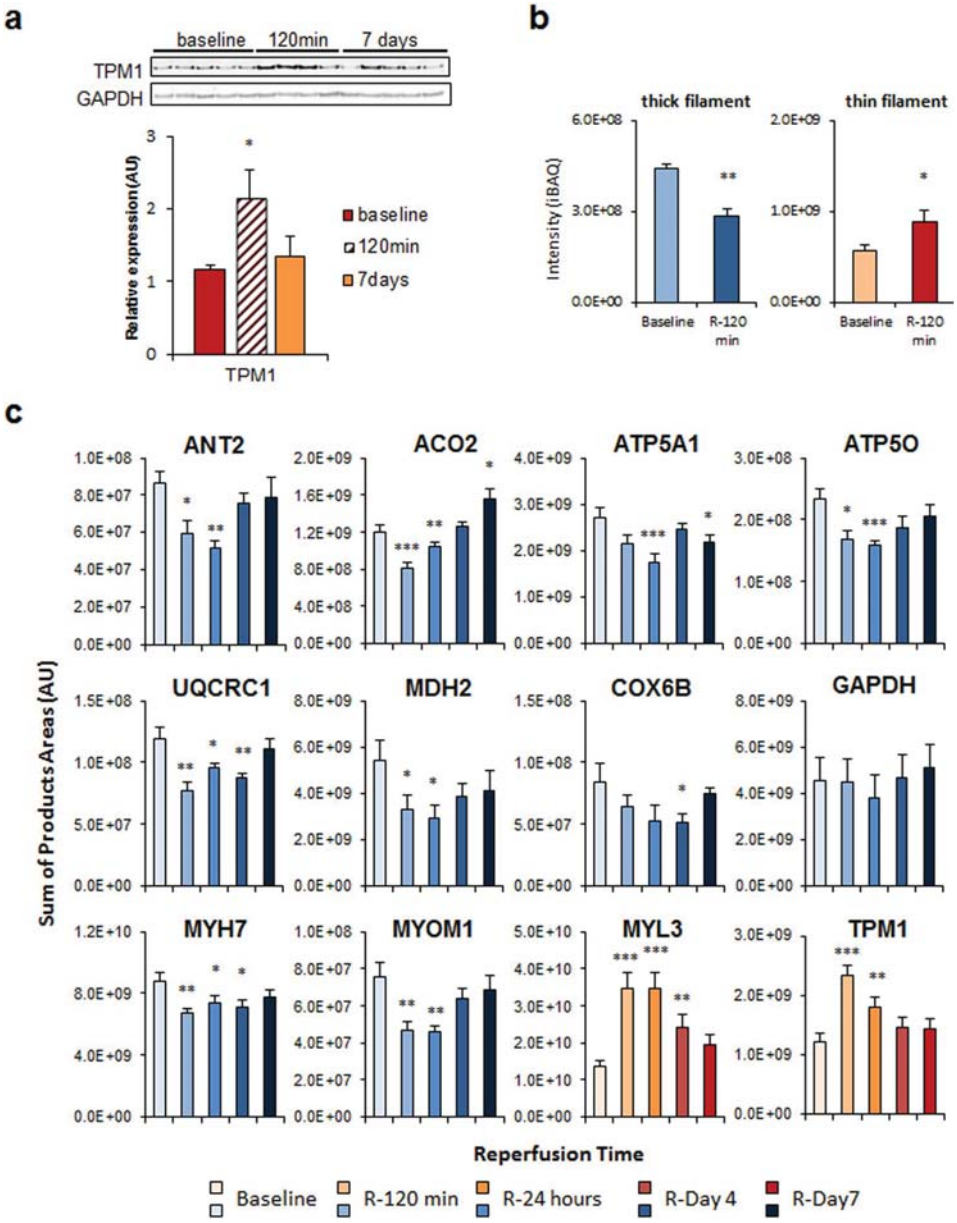


Figure 6. Validation of the most representatitve protein changes detected in remote myocardium after acute infarction. (**a**) Western blot analysis and quantitative densitometry of TPM1 confirming the transient increase in thin filament sarcomeric proteins early after reperfusion. Full-length blots are presented in Supplementary Figure S8 as well as their raw integrated densities before and after normalization can be consulted in Supplementary Figure S12. Protein values are normalized relative to GAPDH. Results are shown as the mean ± s.e.m. of values obtained in at least four animals. *Indicates p value < 0.05 versus baseline by Mann-Whitney test. (**b**) Label free proteomics analysis confirming the decrease in thick filament proteins and an increase in thin filament proteins early after reperfusion (in at least four animals). Results are expressed as in Fig. 4d. Label free individual protein quantifications included in the thick filament and thin filament protein groups are shown in Supplementary Figure S14. (**c**) Quantitative analysis by PRM targeted proteomics of proteins ANT2, ACO2, ATP5A1, ATP5O, UQCRC1, MDH2, COX6B, GAPDH, MYH7, MYOM1, MYL3 and TPM1, confirming the transient changes in abundance of mitochondrial and contractile sarcomeric proteins after early reperfusion. Protein values are normalized by the total base peak area of the chromatograms. Results are shown as the mean ± s.e.m. of at least four animals at each time point in two technical replicates. *Indicates *p value* < 0.05; **indicates *p value* < 0.01, ***indicates *p value* < 0.001 versus baseline by Mann-Whitney test. iTRAQ quantifications of the proteins from panels a and c are shown in the Supplementary Figure S11b for comparison.

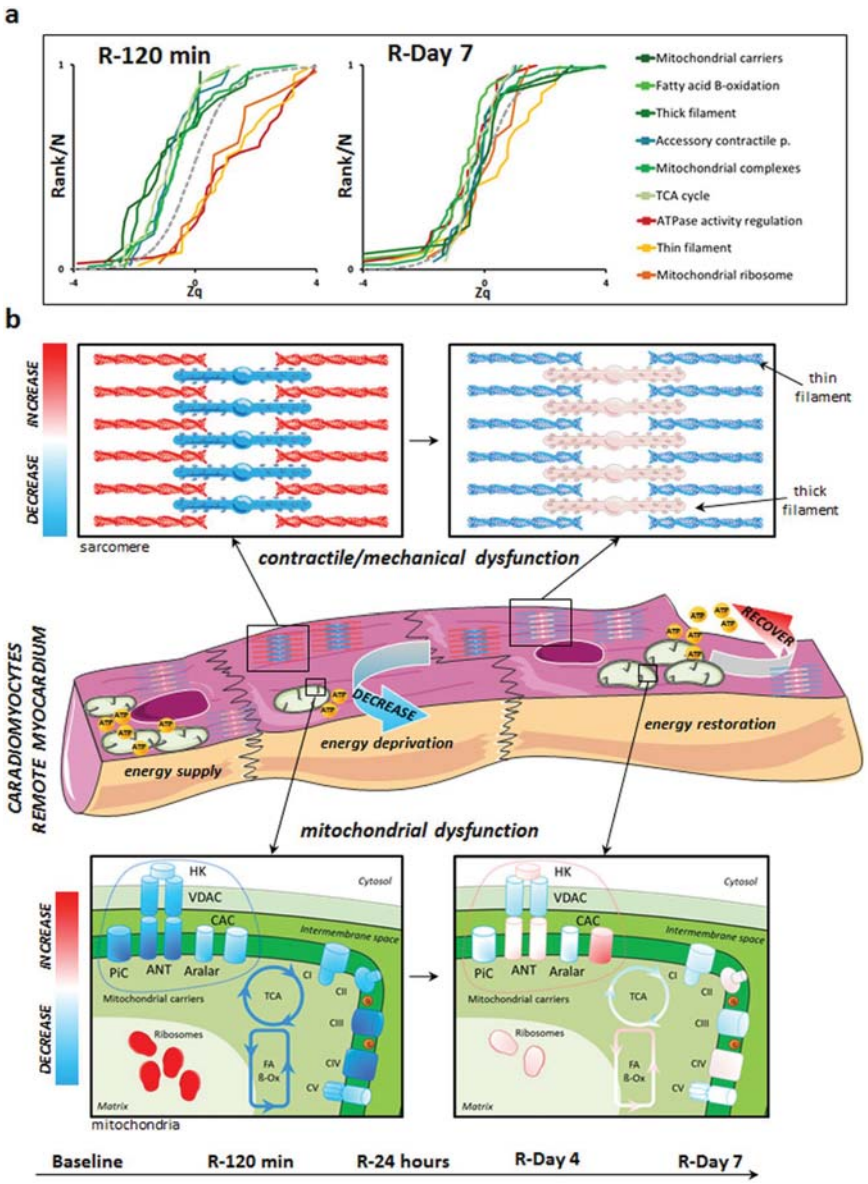


Figure 7. Affected functional processes in sarcomeres and mitochondria of the remote myocardium. (**a**) Cumulative distributions of all Zq values from proteins in the categories depicted in (Supplementary Figure S3) after 120 min (left) or 7 days (right) of reperfusion; for clarity, all proteins belonging to mitochondrial electron transport chain (ETC) complexes are grouped into one category. (**b**) Outline of the mechanical uncoupling within the contractile machinery and the transient decrease in processes crucial for energy production (tricarboxylic acid (TCA) cycle, fatty-acid beta-oxidation, mitochondrial complexes III and IV) and energy transfer and supply (mitochondrial substrate/solute carrier proteins). For clarity, mitochondrial complexes are drawn as simple entities, neglecting their complex structures. Graphical elements in Fig. 7 were adapted from the Servier Medical Art Powerpoint image bank. Servier Medical Art by Servier (http://www.servier.com/Powerpoint-image-bank) is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

molecular alterations of sarcomere components are detected in the remote myocardium, associated with alterations in mitochondrial components (Figs 7a and S3), coinciding with CMR-detected temporal alterations in systolic wall thickness. At later phases (one week post I/R), these changes reverted to the levels detected in the control sample, coinciding with restored systolic contractile function in the remote myocardium.

Acute loss of myocardial function after MI has classically been considered to trigger rapid increases in loading conditions and adaptive responses to preserve stroke volume; these responses include deformation of the remote myocardium, which alters Frank/Starling relations, and activation of the sympathetic adrenergic and renin-angiotensin-aldosterone systems[29]. Our results suggest that these responses might be temporarily maladaptive, because the decrease in regional contractility in the remote myocardium early after reperfusion is almost fully recovered some days later.

Our findings indicate that abrupt post-MI overload of the heart's contractile apparatus and the temporary energy-supply cut-off during early reperfusion induces a compensatory response in the remote myocardium involving transient molecular changes in thin filament components, mitochondrial biosynthesis machinery, and ATPase regulators. However, these adaptive changes are accompanied by a temporary mitochondrial dysfunction that affects energy-producing components (ETC, TCA cycle, fatty acid beta-oxidation, and mitochondrial energy carriers), while there are transient alterations in other sarcomere components (thick filaments and accessory contractile proteins) (Fig. 7b). This view is supported by recent evidence that thin and thick filaments are independently and separately regulated[30]. Likewise, transcriptional analysis of the failing mammalian myocardium suggests an orchestrated dysregulation of energy supply (metabolism) and expenditure (muscle contraction and ion homeostasis)[31]. In addition, the strong alteration of mitochondrial complexes III and IV (at 120 min and 24 h, Supplementary Figure S2) is in good agreement with a recent report that succinate accumulating during ischemia in mice is rapidly metabolized after reperfusion, producing a large proton motive force through ETC complexes III and IV[32].

State-of-the-art T2 mapping CMR sequences show a moderate edematous reaction appearing progressively several days after I/R, evidenced by the progressive increase in T2 relaxation times during the first week after reperfusion[18,22]. These results also correlate precisely with the proteomics analysis in the remote myocardium, which shows changes in proteins involved in inflammation, wound healing, and fibrotic processes. Notably, edema in the remote region follows a different pattern to that found in the ischemic region[18,22]. The at-risk region shows a bimodal pattern of edema, with the initial wave secondary to reperfusion and the deferred associated with tissue repair; in contrast, in the remote region edema formation is unimodal and progressive. In the infarcted region, activation of inflammatory response proteins coincides with the initial edema wave, whereas in the remote region it correlates with the later progressive edema formation. This observation reinforces the idea that these edematous processes are closely related to inflammation and indicate that the inflammatory response in the remote region is delayed by some days relative to the reaction in the infarcted region. Inflammation also coincided with increased collagen levels that were also observed in the remote myocardium at these later reperfusion phases[10]. A subtle post-MI inflammation in the remote myocardium is also supported by the accumulation in this region of monocytes and macrophages, albeit more slowly and in lower amounts than the ischemic myocardium[33].

The different timing of inflammatory responses in the remote and ischemic post-MI myocardium might play an important role in modulating extracellular matrix protein composition, which contributes to adverse remodeling of the remote region and heart failure[33–37]. Well-healed infarcts contain large amounts of extracellular matrix, occupying up to 80% of the infarcted area. Collagen deposition in the non-infarcted remote myocardium occurs predominantly in the interstitium, where it contributes to ventricular stiffness and dysfunction[16]. Although collagen deposition initially helps to maintain ventricular morphology after MI, over time it can contribute to geometric changes and functional deterioration[35]. Phenotypically transformed myofibroblasts and inflammatory cells, including those located in the remote myocardium, seem to be central to fibrosis at remote sites after MI due to their possible contribution to the upregulation of the renin-angiotensin system, which is strongly implicated in ventricular remodeling[16,17,38].

Limitations. The results presented are from a juvenile healthy animal model in the absence of comorbidities and classical risk factors for cardiovascular disease; therefore, care should be taken in extrapolating to the clinic[39]. Nonetheless, the pig is one of the most clinically translatable large animal models for the study of I/R issues, because unlike other mammals its coronary artery anatomy and distribution resembles that of humans, including a minimal pre-existing coronary collateral flow[40]. The use of a large animal model is of great translational value, especially considering the difficulty of performing such a comprehensive myocardial tissue analysis at the molecular level in patients. The time course of tissue changes in the post-I/R myocardium in pigs, albeit slightly briefer, is similar to that observed in humans[10,41]. In general, cardiac contractility and performance, and therefore cardiac tolerance against ischemia/reperfusion injury might be influenced by anesthesia with additional potential inter- and intra-individual variations in experimental models. We tried to overcome this limitation by performing different but complimentary analysis as follows. Firstly, we analyzed the individual percentage SWT change relative to baseline; therefore taking account for inter-individual variations. Secondly, we analyzed SWT results by using generalized linear mixed models for ischemic and remote myocardium separately; therefore taking account for repeated measurements, inter- and intra-individual variations. Finally, we demonstrated that infarct transmurality and final tissue damage was very similar among pigs. All these data together reinforce the notion that the time profile of regional contractility and final injury inflicted after I/R was quite homogeneous in our experimental model with little influence (if any) of anesthesia in our results, though we cannot definitely exclude any potential interaction.

Another limitation is that the proteomics results in the present study come from homogenized myocardial tissue; therefore it is not possible to differentiate protein expression in different cell types or myocardial layers. Similarly, our data lacks information on subcellular location, on post-translational modifications and on the activity of the proteins quantified. A related problem is the possibility that during reperfusion, blood may occasionally enter the interstitial area through damaged vessels, influencing the proteomics results. The total amount of proteins potentially originating from plasma is, however, very low (Supplementary Figure S7), suggesting that if present, this kind of contamination is insufficient to significantly alter our results.

Finally, the incomplete annotation and functional classification of *Sus scrofa* proteins is also a problem that difficulties biological interpretation. Although we have managed this issue by assigning pig proteins to their human counterparts on the basis of shared peptide sequences, extrapolation of pig protein results to humans should be done with caution.

We present the first comprehensive proteomics analysis of the post-I/R myocardium, covering the first week after infarction. Our results highlight a dynamic pattern of molecular responses that take place in the ischemic and remote myocardium, linked to functional changes revealed by CMR. The results provide evidence of intriguing transformations in the remote non-infarcted myocardium during the first week after I/R, as demonstrated by striking local proteome changes, temporarily reduced contractility function, and the presence of edema in this region. These findings provide insight into the precise molecular signatures underlying early remodeling stages in both the ischemic and the remote myocardium, with potential for the development of new therapeutic approaches to preventing post-MI HF.

Methods

Study design and myocardial infarction procedure. The study population was formed by 25 castrated male 3 to 4 month old Large-White pigs weighing 30 to 40 kg. The studies on swine were conducted in accordance with the *Guide for the Care and Use of Laboratory Animals* (the Institute of Laboratory Animal Resources, 1996), and with the approval of Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC) Institutional Animal Research Committee (CNIC 05/13) and the Regional Animal Research Committee (PROEX 51/13). We previously reported the dynamic changes in edema formation (assessed by T2 mapping CMR) and tissue composition (assessed by histology) in the ischemic area in same cohort of pigs[18]. The study design is summarized in Fig. 1a. Five pigs (Group 1) were sacrificed with no intervention other than baseline CMR, and served as controls. In 20 pigs, reperfused transmural acute myocardial infarction (I/R) was induced experimentally by closed-chest 40-minute mid left anterior descending coronary artery occlusion followed by balloon deflation and reestablishment of blood flow[18,41,42]. These pigs were sacrificed at 120 min (n = 5, Group 2), 24 h (n = 5, Group 3), 4 days (n = 5, Group 4), and 7 days (n = 5, Group 5). CMR scans were performed at every follow-up stage until sacrifice (i.e. animals sacrificed on day 7 underwent baseline, 120 min, 24 h, day 4, and day 7 CMR exams). The animals were euthanized immediately after the last follow-up CMR scan, and transmural myocardial tissue samples from ischemic and remote areas were rapidly collected for proteomic evaluations. Based on anatomical correlates and standard left ventricle segmentation, those areas from mid-apical ventricular short axis slices matching regional contractility analysis were selected for sampling collection. Detailed information about the study design and myocardial infarction procedure can be found in a previous publication[18].

CMR protocol and analysis. CMR exams were performed immediately before MI induction and at subsequent post-MI follow-up time points until sacrifice. All CMR studies were conducted with a Philips 3-Tesla Achieva Tx whole body scanner (Philips Healthcare, Best, the Netherlands) equipped with a 32-element phased-array cardiac coil. The imaging protocol included a standard segmented cine steady-state free-precession (SSFP) sequence to provide high quality anatomical references and segmental regional contractility[18,41,42]. CMR images were analyzed using dedicated software for regional contractility (QMass MR 7.6; Medis, Leiden, the Netherlands) by two observers experienced in CMR analysis and blinded to group allocation. To quantify end-diastolic wall thickness (EDWT), end-systolic wall thickness (ESWT), LV volumes, and LV mass, epicardial and endocardial contours were detected automatically, and corrected manually on short-axis cine SSFP if needed. Systolic wall thickening (SWT) was calculated according to the formula (ESWT − EDWT)/EDWT ×100. SWT was summed and then averaged from mid-apical anteroseptal and anterior segments, for the ischemic area, and mid-apical inferolateral and inferior segments, for the remote area. Myocardial segments were considered dysfunctional if SWT was 30% or lower[43]. Detailed information about imaging parameters and CMR analysis can be found in the Supplementary Information file.

Mass spectrometry analysis, and protein identification and quantification. Pigs were sacrificed at different time-points after CMR (120 min, 1 day, 4 days, and 7 days post-reperfusion). Samples from the ischemic and remote myocardium of all pigs were collected within minutes of euthanasia and processed for proteomic analysis (Fig. 1b). The 40 minutes-I/R protocol applied in our experimental setting is able to induce a transmural necrosis with virtually no viable myocardium within the ischemic area[10,18,41,42]; therefore ensuring analysis at pure infarcted and remote myocardium levels. Protein extracts were obtained by tissue homogenization with ceramic beads (MagNa Lyser Green Beads apparatus, Roche, Germany) in extraction buffer (50 mM Tris-HCl, 1 mM EDTA, 1.5% SDS, pH 8.5). Free Cys residues were blocked with 50 mM iodoacetamide at the time of protein extraction. For the screening proteomics approach, the protein extracts for each time-point from five biological replicate samples were pooled according to their concentration. Samples were subjected to tryptic digestion and the resulting peptides were labeled with 8-plex isobaric tags for relative and absolute quantification (iTRAQ) and separated by cation exchange chromatography. The fractionated peptides were analyzed by nano-liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) using a Q-Exactive hybrid quadrupole orbitrap mass spectrometer (Thermo Scientific). The proteomics results presented in Figs 3 and 5 are representative of three technical replicate experiments.

Protein identification was performed using the SEQUEST HT algorithm integrated in Proteome Discoverer 1.4 (Thermo Scientific). Since *Sus scrofa* gene and protein annotation is not complete, MS/MS scans were searched against a combined pig and human database (UniProtKB/Swiss-Prot 2014_02 Release); pig proteins are given priority when they share peptides with human proteins. Peptides were identified from MS/MS data using the probability ratio method[44]. False discovery rate (FDR) of peptide identifications was calculated by the

refined method[45,46]. Quantitative information was extracted from the MS/MS spectra of iTRAQ-labeled peptides using the in-house program QuiXoT, as described[47]. Differential protein expression was analyzed using the WSPP model[47,48], which uses raw quantifications as input data and computes the protein log_2-fold changes (FC) expressed in units of standard deviation around the averages (Zq) for each condition (Groups 2–5), with respect to the control (Group 1). For protein quantification, no limit was imposed on the minimum number of peptides per protein[47,48]. A quantitative proteomics result of total protein changes globally between remote and ischemic myocardium during the first week of reperfusion is summarized in Supplementary Table S10. Alterations in biological functions as a consequence of the coordinated behavior of proteins, was analyzed by estimating functional category averages (Zc) according to the Systems Biology Triangle (SBT) model[19]. The quantified proteins were functionally annotated using the Ingenuity Knowledge Database (IPA)[49,50] and DAVID[51]. The DAVID repository includes 13 functional databases, including Gene Ontology, KEGG, and Panther. Further details about proteomics and statistical analysis can be found in the Supplementary Information file.

Parallel Reaction Monitoring Analysis. Samples from individual pigs used in this analysis were obtained from five animals for each time-point (baseline, R-24 hours, R-Day 4 and R-Day 7) and four animals for time-point R-120 min. For the targeted monitoring, we selected the peptides which were identified with the highest SEQUEST score (Xcorr) and that belonged to representative proteins from the functional categories (Supplementary Table S11). Quantitative targeted protein analyses for each biological replicates were performed by parallel reaction monitoring (PRM)[52] by nano-liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) using a QExactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) in two technical replicates. Raw mass spectrometry files (Thermo) and spectral libraries (msf files obtained with Proteome Discoverer 2.1 search engine) were imported into Skyline version 3.6.0.10162 ((https://skyline.gs.washington.edu) for identification of transitions and peak area integration according to the software instructions[53]. Only b- or y- fragment ions were selected to build the elution profile for peptide quantitation. All extracted ion chromatograms (XICs) of selected fragments were manually inspected and adjusted to ensure proper peak picking and peak integration.

Label Free Proteomics Analysis. Samples used in this analysis obtained from 4 individual pigs (time-point R-120 min) or 5 individual pigs (time-points baseline, R-24 hours, R-Day 4 and R-Day 7) were processed as described above. Label free experiments for all biological replicates were acquired using nano-liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) in a QExactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific). MS data were acquired with a Top10 data-dependent MS/MS scan method (topN method). Mass spectrometry raw files were analyzed by MaxQuant software (version 1.5.6.5)[54]. Label-free protein quantitation (LFQ) was performed with a minimum ratio count of 1[55]. Intensity-based absolute quantification (iBAQ)[20] implemented in MaxQuant was used for the analysis of the label-free proteomics experiments. We express the protein abundances as percentage of the identified proteome, obtained by normalizing the iBAQ intensities to the sum of all intensities.

Western Immunoblotting Analysis. Validation by western blot was performed on selected representative proteins from each functional category according to antibody availability, giving priority to the proteins identified with the highest number of peptides by mass spectrometry. Immunoblotting was performed according to standard protocols. Briefly, 10 μg of heart tissue extracts were loaded per lane separated on the 4–10% gel and transferred to polyvinylidene fluoride (PVDF) (Immobilon-FL, Milipore) membranes for fluorescence applications. Validation of proteomics results was performed using the following antibodies: ATP5A (Abcam, ab14748), UQCRC2 (Abcam, ab14745), SDHB (Abcam, ab14714), NDUFB8 (Abcam, ab110242), UQCRFS1 (Abcam, ab14746), COXIV (Abcam, ab14744), ATP5B (Abcam, ab14730), cardiac Myosin (Abcam, ab50967), TPM1 (Abcam, ab133292), TNNT2 (Abcam, ab10214) and GAPDH (Santa Cruz Biotechnology, sc-32233). Secondary antibodies goat anti-mouse IgG DyLight 800 (Rockland, 610-145-121) and goat anti-rabbit IgG Alexa Fluor 680 (Thermo Fisher Scientific, A-21076) were used against the corresponding primary antibodies and the images were acquired with the ODYSSEY Infrared Imaging System (LI-COR).

Statistical analysis. Normal distribution of each data subset was checked using graphical methods and a Shapiro–Wilk test. For quantitative CMR variables, data are expressed as mean and standard deviation. To take into account repeated measures, the time course of regional contractility was analyzed with generalized linear mixed models. Pairwise comparisons were made between baseline and all other CMR time points, and p-values were adjusted for multiple comparisons by the sequential Holm-Bonferroni method. Statistical analyses were performed with Stata 12.0 (StataCorp, College Station, Texas). Western blot analysis to determine significant differences between groups after densitometry analysis and statistical analysis of absolute peptide quantitation by PRM were performed using the Mann-Whitney test, using the GraphPad Prism 7.02 software. Results of intensity-based absolute quantification (iBAQ) of label free proteomics experiments were tested for significance using the Mann-Whitney test with a false discovery rate of <0.05 after Benjamini–Hochberg correction.

Data availability. The dataset from the analysis of the pig myocardium proteome (raw files and excel tables with peptide and protein quantification data), the label-free and the PRM analysis raw data as well as the raw western blot data from the ODISSEY scanner are available in the PeptideAtlas repository http://www.peptideatlas.org/PASS/PASS00891), which can be downloaded via ftp.peptideatlas.org.

References

1. Menees, D. S. *et al*. Door-to-balloon time and mortality among patients undergoing primary PCI. *The New England journal of medicine* **369**, 901–909, https://doi.org/10.1056/NEJMoa1208200 (2013).
2. Eapen, Z. J. *et al*. Defining heart failure end points in ST-segment elevation myocardial infarction trials: integrating past experiences to chart a path forward. *Circulation. Cardiovascular quality and outcomes* **5**, 594–600, https://doi.org/10.1161/CIRCOUTCOMES.112.966150 (2012).
3. de Couto, G., Ouzounian, M. & Liu, P. P. Early detection of myocardial dysfunction and heart failure. *Nat Rev Cardiol* **7**, 334–344, https://doi.org/10.1038/nrcardio.2010.51 (2010).
4. Lindsey, M. L. *et al*. Transformative Impact of Proteomics on Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association. *Circulation* **132**, 852–872, https://doi.org/10.1161/CIR.0000000000000226 (2015).
5. Varga, Z. V. *et al*. Functional Genomics of Cardioprotection by Ischemic Conditioning and the Influence of Comorbid Conditions: Implications in Target Identification. *Curr Drug Targets* **16**, 904–911 (2015).
6. Chen, W. *et al*. Endothelial Actions of ANP Enhance Myocardial Inflammatory Infiltration in the Early Phase After Acute Infarction. *Circulation research*. https://doi.org/10.1161/CIRCRESAHA.115.307196 (2016).
7. Kain, V., Prabhu, S. D. & Halade, G. V. Inflammation revisited: inflammation versus resolution of inflammation following myocardial infarction. *Basic Res Cardiol* **109**, 444, https://doi.org/10.1007/s00395-014-0444-7 (2014).
8. Peng, Y. *et al*. Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. *Mol Cell Proteomics* **13**, 2752–2764, https://doi.org/10.1074/mcp.M114.040675 (2014).
9. Basso, C., Rizzo, S. & Thiene, G. The metamorphosis of myocardial infarction following coronary recanalization. *Cardiovascular pathology: the official journal of the Society for Cardiovascular Pathology* **19**, 22–28, https://doi.org/10.1016/j.carpath.2009.06.010 (2010).
10. Fernandez-Jimenez, R. *et al*. Pathophysiology Underlying the Bimodal Edema Phenomenon After Myocardial Ischemia/Reperfusion. *J Am Coll Cardiol* **66**, 816–828, https://doi.org/10.1016/j.jacc.2015.06.023 (2015).
11. Fernandez-Caggiano, M. *et al*. Analysis of Mitochondrial Proteins in the Surviving Myocardium after Ischemia Identifies Mitochondrial Pyruvate Carrier Expression as Possible Mediator of Tissue Viability. *Molecular & cellular proteomics: MCP* **15**, 246–255, https://doi.org/10.1074/mcp.M115.051862 (2016).
12. Barallobre-Barreiro, J. *et al*. Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury. *Circulation* **125**, 789–802, https://doi.org/10.1161/CIRCULATIONAHA.111.056952 (2012).
13. Chang, Y. H. *et al*. Quantitative proteomics reveals differential regulation of protein expression in recipient myocardium after trilineage cardiovascular cell transplantation. *Proteomics* **15**, 2560–2567, https://doi.org/10.1002/pmic.201500131 (2015).
14. Cieniewski-Bernard, C. *et al*. Proteomic analysis of left ventricular remodeling in an experimental model of heart failure. *J Proteome Res* **7**, 5004–5016, https://doi.org/10.1021/pr800409u (2008).
15. Page, B. J. *et al*. Revascularization of chronic hibernating myocardium stimulates myocyte proliferation and partially reverses chronic adaptations to ischemia. *J Am Coll Cardiol* **65**, 684–697, https://doi.org/10.1016/j.jacc.2014.11.040 (2015).
16. van den Borne, S. W. *et al*. Myocardial remodeling after infarction: the role of myofibroblasts. *Nat Rev Cardiol* **7**, 30–37, https://doi.org/10.1038/nrcardio.2009.199 (2010).
17. Heusch, G. *et al*. Cardiovascular remodelling in coronary artery disease and heart failure. *The Lancet* **383**, 1933–1943, https://doi.org/10.1016/s0140-6736(14)60107-0 (2014).
18. Fernandez-Jimenez, R. *et al*. Myocardial edema after ischemia/reperfusion is not stable and follows a bimodal pattern: imaging and histological tissue characterization. *J Am Coll Cardiol* **65**, 315–323, https://doi.org/10.1016/j.jacc.2014.11.004 (2015).
19. Garcia-Marques, F. *et al*. A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics. *Mol Cell Proteomics* **15**, 1740–1760, https://doi.org/10.1074/mcp.M115.055905 (2016).
20. Schwanhausser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342, https://doi.org/10.1038/nature10098 (2011).
21. Zamilpa, R. & Lindsey, M. L. Extracellular matrix turnover and signaling during cardiac remodeling following MI: causes and consequences. *J Mol Cell Cardiol* **48**, 558–563, https://doi.org/10.1016/j.yjmcc.2009.06.012 (2010).
22. Fernandez-Jimenez, R. *et al*. Fast T2 gradient-spin-echo (T2-GraSE) mapping for myocardial edema quantification: first *in vivo* validation in a porcine model of ischemia/reperfusion. *J Cardiovasc Magn Reson* **17**, 92, https://doi.org/10.1186/s12968-015-0199-9 (2015).
23. Ibanez, B., Heusch, G. & Ovize, M. & Van de Werf, F. Evolving therapies for myocardial ischemia/reperfusion injury. *J Am Coll Cardiol* **65**, 1454–1471, https://doi.org/10.1016/j.jacc.2015.02.032 (2015).
24. Vogel, B., Shinagawa, H., Hofmann, U., Ertl, G. & Frantz, S. Acute DNase1 treatment improves left ventricular remodeling after myocardial infarction by disruption of free chromatin. *Basic Res Cardiol* **110**, 15, https://doi.org/10.1007/s00395-015-0472-y (2015).
25. Ong, S. B., Samangouei, P., Kalkhoran, S. B. & Hausenloy, D. J. The mitochondrial permeability transition pore and its role in myocardial ischemia reperfusion injury. *Journal of molecular and cellular cardiology* **78**, 23–34, https://doi.org/10.1016/j.yjmcc.2014.11.005 (2015).
26. Inserte, J. *et al*. Studies on the role of apoptosis after transient myocardial ischemia: genetic deletion of the executioner caspases-3 and -7 does not limit infarct size and ventricular remodeling. *Basic Res Cardiol* **111**, 18, https://doi.org/10.1007/s00395-016-0537-6 (2016).
27. Smeele, K. M. *et al*. Disruption of hexokinase II-mitochondrial binding blocks ischemic preconditioning and causes rapid cardiac necrosis. *Circulation research* **108**, 1165–1169, https://doi.org/10.1161/CIRCRESAHA.111.244962 (2011).
28. Chan, W. *et al*. Acute left ventricular remodeling following myocardial infarction: coupling of regional healing with remote extracellular matrix expansion. *JACC Cardiovasc Imaging* **5**, 884–893, https://doi.org/10.1016/j.jcmg.2012.03.015 (2012).
29. Sutton, M. G. & Sharpe, N. Left ventricular remodeling after myocardial infarction: pathophysiology and therapy. *Circulation* **101**, 2981–2988 (2000).
30. Kampourakis, T., Sun, Y. B. & Irving, M. Myosin light chain phosphorylation enhances contraction of heart muscle via structural changes in both thick and thin filaments. *Proc Natl Acad Sci USA* **113**, E3039–3047, https://doi.org/10.1073/pnas.1602776113 (2016).
31. Barth, A. S., Kumordzie, A. & Tomaselli, G. F. Orchestrated regulation of energy supply and energy expenditure: Transcriptional coexpression of metabolism, ion homeostasis, and sarcomeric genes in mammalian myocardium. *Heart Rhythm* **13**, 1131–1139, https://doi.org/10.1016/j.hrthm.2016.01.009 (2016).
32. Chouchani, E. T. *et al*. Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* **515**, 431–435, https://doi.org/10.1038/nature13909 (2014).
33. Frantz, S. & Nahrendorf, M. Cardiac macrophages and their role in ischaemic heart disease. *Cardiovasc Res* **102**, 240–248, https://doi.org/10.1093/cvr/cvu025 (2014).
34. Hori, M. & Nishida, K. Oxidative stress and left ventricular remodelling after myocardial infarction. *Cardiovasc Res* **81**, 457–464, https://doi.org/10.1093/cvr/cvn335 (2009).
35. Baysa, A. *et al*. The p66ShcA adaptor protein regulates healing after myocardial infarction. *Basic Res Cardiol* **110**, 13, https://doi.org/10.1007/s00395-015-0470-0 (2015).
36. Martire, A. *et al*. Mesenchymal stem cells attenuate inflammatory processes in the heart and lung via inhibition of TNF signaling. *Basic Res Cardiol* **111**, 54, https://doi.org/10.1007/s00395-016-0573-2 (2016).
37. Robinson, E. *et al*. Exendin-4 protects against post-myocardial infarction remodelling via specific actions on inflammation and the extracellular matrix. *Basic Res Cardiol* **110**, 20, https://doi.org/10.1007/s00395-015-0476-7 (2015).

38. Sun, Y. Myocardial repair/remodelling following infarction: roles of local factors. *Cardiovasc Res* **81**, 482–490, https://doi.org/10.1093/cvr/cvn333 (2009).
39. Ferdinandy, P., Hausenloy, D. J., Heusch, G., Baxter, G. F. & Schulz, R. Interaction of risk factors, comorbidities, and comedications with ischemia/reperfusion injury and cardioprotection by preconditioning, postconditioning, and remote conditioning. *Pharmacol Rev* **66**, 1142–1174, https://doi.org/10.1124/pr.113.008300 (2014).
40. Fernández-Jiménez, R., Fernández-Friera, L., Sánchez-González, J. & Ibáñez, B. Animal Models of Tissue Characterization of Area at Risk, Edema and Fibrosis. *Curr Cardiovasc Imaging Rep* **7**, 1–10, https://doi.org/10.1007/s12410-014-9259-z (2014).
41. Fernández-Jiménez, R. *et al*. Dynamic Edematous Response of the Human Heart to Myocardial Infarction: Implications for Assessing Myocardial Area at Risk and Salvage. *Circulation*, doi:https://doi.org/10.1161/circulationaha.116.025582 (2017).
42. Fernández-Jiménez, R. *et al*. Effect of Ischemia Duration and Protective Interventions on the Temporal Dynamics of Tissue Composition After Myocardial Infarction. *Circ Res* **121**, 439–450, https://doi.org/10.1161/CIRCRESAHA.117.310901 (2017).
43. Nowosielski, M. *et al*. Comparison of wall thickening and ejection fraction by cardiovascular magnetic resonance and echocardiography in acute myocardial infarction. *Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance* **11**, 22, https://doi.org/10.1186/1532-429X-11-22 (2009).
44. Martinez-Bartolome, S. *et al*. Properties of average score distributions of SEQUEST: the probability ratio method. *Mol Cell Proteomics* **7**, 1135–1145, https://doi.org/10.1074/mcp.M700239-MCP200 (2008).
45. Navarro, P. & Vazquez, J. A refined method to calculate false discovery rates for peptide identification using decoy databases. *J Proteome Res* **8**, 1792–1796, https://doi.org/10.1021/pr800362h (2009).
46. Bonzon-Kulichenko, E., Garcia-Marques, F., Trevisan-Herraz, M. & Vazquez, J. Revisiting peptide identification by high-accuracy mass spectrometry: problems associated with the use of narrow mass precursor windows. *J Proteome Res* **14**, 700–710, https://doi.org/10.1021/pr5007284 (2015).
47. Navarro, P. *et al*. General statistical framework for quantitative proteomics by stable isotope labeling. *J Proteome Res* **13**, 1234–1247, https://doi.org/10.1021/pr4006958 (2014).
48. Martinez-Acedo, P. *et al*. A novel strategy for global analysis of the dynamic thiol redox proteome. *Molecular & cellular proteomics: MCP* **11**, 800–813, https://doi.org/10.1074/mcp.M111.016469 (2012).
49. Calvano, S. E. *et al*. A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–1037, https://doi.org/10.1038/nature03985 (2005).
50. Ficenec, D. *et al*. Computational knowledge integration in biopharmaceutical research. *Brief Bioinform* **4**, 260–278 (2003).
51. Huang da, W. *et al*. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* Chapter 13, Unit 13 11, doi:https://doi.org/10.1002/0471250953.bi1311s27 (2009).
52. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **11**, 1475–1488, https://doi.org/10.1074/mcp.O112.020131 (2012).
53. MacLean, B. *et al*. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968, https://doi.org/10.1093/bioinformatics/btq054 (2010).
54. Cox, J. & Mann, M. In *Nat Biotechnol* Vol. 26 1367–1372 (2008).
55. Cox, J. *et al*. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513–2526, https://doi.org/10.1074/mcp.M113.031591 (2014).

Acknowledgements

We thank Tamara Córdoba, Oscar Sanz, Eugenio Fernández and other members of the CNIC animal facility and farm for outstanding animal care and support. Simon Bartlett (CNIC) provided English editing. Graphical elements in Figures 1 and 7 were adapted from the Servier Medical Art Powerpoint image bank. Servier Medical Art by Servier (http://www.servier.com/Powerpoint-image-bank) is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/). This study was supported by competitive grants from the Spanish Ministry of Economy and Competitiveness (MINECO) (BIO2015-67580-P) through the Carlos III Institute of Health-Fondo de Investigación Sanitaria (PI13/01979 and PRB2 (IPT13/0001 - ISCIII-SGEFI/FEDER, ProteoRed)), the Fondo Europeo de Desarrollo Regional (FEDER, RD: SAF2013-49663-EXP), the Fundación La Marato TV3 and in part by the FP7-PEOPLE-2013-ITN *Next generation training in cardiovascular research and innovation-Cardionext*. Borja Ibanez and Jesús Vázquez are supported by the *Red de Investigación Cardiovascular* (RIC) of the Spanish Ministry of Health (RD 12/0042/0054 and RD12/0042/0056). Rodrigo Fernández-Jiménez is a recipient of non-overlapping grants from the Ministry of Economy and Competitiveness through the *Instituto de Salud Carlos III* (*Rio Hortega* fellowship); and the *Fundació Jesús Serra*, the *Fundación Interhospitalaria de Investigación Cardiovascular* (FIC), and the CNIC (FICNIC Fellowship). Jaume Agüero, Aleksandra Binek, and Navratan Bagwan are FP7-PEOPLE-2013-ITN-Cardionext fellows. This work has been co-funded by FEDER funds. This study forms part of a Master Research Agreement (MRA) between CNIC and Philips Healthcare. The CNIC is supported by the MINECO and the Pro-CNIC Foundation, and is a Severo Ochoa Center of Excellence (MINECO award SEV-2015–0505).

Author Contributions

A.B. and R.F.J. designed and planned all the experiments, performed statistical analysis, and drafted the manuscript. C.G.A. and J.A. performed the infarct procedure in the pig and helped with CMR imaging analysis. N.B. helped with the bioinformatics analysis of proteomics data. A.P. helped with the validation experiments by western blot. I.J. and E.C. helped with the proteomics validation experiments. E.C., J.A.L. and V.F. contributed with interpretation of data for the work and scientific support. I. J., J.V. and B.I. designed and supervised the experiments, corrected the manuscript, coordinated the project, and handled funding. All authors revised the manuscript for significant intellectual content and gave final approval of the version to be published.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-11985-5.

Competing Interests: The authors declare that they have no competing interests.

4. Other paper published during PhD

4.1. Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators

Analytica Chimica Acta xxx (2018) 1–9

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators

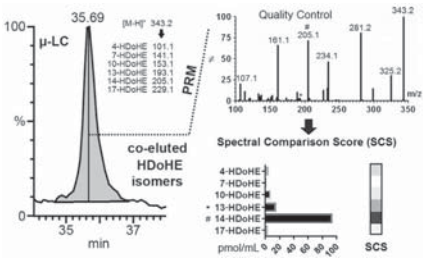
Adam Wutkowski [a, 1], Matthias Krajewski [a, 1], Navratan Bagwan [b, 1], Mathias Schäfer [c], Bhesh R. Paudyal [d], Ulrich E. Schaible [d], Dominik Schwudke [a, *]

[a] Division of Bioanalytical Chemistry, Research Center Borstel, Parkallee 10, Borstel, Germany
[b] Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Melchor Fernandez Almagro 3, 28029 Madrid, Spain
[c] Institute of Organic Chemistry, University of Cologne, Greinstraße 4, 50939 Köln, Germany
[d] Department of Cellular Microbiology, Research Center Borstel, Parkallee 10, Borstel, Germany

HIGHLIGHTS

- LC-MS^2 method for lipid mediator quantitation compatible with SPE-free lipid extraction is characterized.
- Resource of tandem mass spectra of lipid mediators is provided.
- CID MS^2 of lipid mediators can comprise odd-electron fragment ions.
- Spectrum Comparison Score (SCS) provides quality control for PRM-based quantitation.
- Platform independent software SpeCS is introduced to manage customized spectral libraries and compute SCS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 October 2017
Received in revised form
25 January 2018
Accepted 29 January 2018
Available online xxx

Keywords:
Lipid mediators
Liquid Chromatography–Tandem mass spectrometry
Parallel reaction monitoring
Spectral library
Spectrum comparison

ABSTRACT

We characterized the performance of a micro-flow LC-ESI-MS^2 approach to analyze lipid mediators (LMs) and polyunsaturated fatty acids (PUFA) that was optimized for SPE free lipid extraction. Tandem mass spectrometry was exclusively performed in parallel reaction monitoring (PRM) mode using TOF and Orbitrap analyzers. This acquisition strategy allowed in addition to quantitation by specific quantifier ions to perform spectrum comparisons using full MS^2 spectra information of the analyte. Consequently, we developed a dedicated software SpeCS that allows to 1) process raw peak lists, 2) generate customized spectral libraries, 3) test specificity of quantifier ions and 4) perform spectrum comparisons. The dedicated scoring algorithm is based on signal matching and Spearman's rank correlation of intensities of matched signal. The algorithm was evaluated in respect of its specificity to distinguish structural related LMs on both instrument platforms. We show how high resolution mass spectrometry is beneficial to distinguish co-eluted LM isomers and provide a generalized quality control procedure for PRM. The applicability of the approach was evaluated analyzing the lipid mediator response during *M. tuberculosis* infection in the mouse lung.

* Corresponding author.
E-mail address: dschwudke@fz-borstel.de (D. Schwudke).
[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.aca.2018.01.044
0003-2670/

1. Introduction

Lipid mediators (LMs) are derivatives of omega-3 and -6 (ω-3, ω-6) polyunsaturated fatty acids (PUFAs) like arachidonic acid (AA), eicosapentanoic acid (EPA) or docosahexanoic acid (DHA) by enzymatic conversions through lipoxygenases (LOXs) and/or cyclooxygenases (COXs) [1–3]. The oxidation of these C20- and C22-carboxylate bodies generates messengers comprising hundreds of constitutional isomers, which can be released from diverse cellular sources in response to microbial infections, physical and chemical stresses [4]. The resulting LM profile reflects the host's inflammation status and its precise chemical determination can be potentially exploited for diagnostic purposes [5]. Thus, in the initiation phase of inflammation pro-inflammatory LMs, particularly leukotrienes (LTs) and prostaglandins (PGs) are preponderant, while the resolution of inflammation, in contrast, is dominated by pro-resolving LMs, such as lipoxins (LXs), maresin (Mar), protectin (PD) and resolvins of the E- and D-series (RvEs, RvDs) [6]. Herein, all LM types are thought to exert their regulatory effects by binding to G protein-coupled receptors (GPCR), this contributing to the control of all aspects of tissue inflammation, including e.g., the recruitment of inflammatory cells, such as neutrophils, monocytes, and macrophages, into peripheral tissues [7–9]. However, one of the major challenge in study of lipid mediators is the unequivocally identification of isomers that can fulfill diametral roles in homeostatic and pathophysiological events [10,11]. In recent years, mass spectrometry (MS) based methodologies have become the state-of-the-art technology for LMs identification and quantitation in complex biological tissues. Several targeted approaches were described to analyze oxidized PUFAs on the basis of liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS^2) [12–17]. Generally, LC-MS^2 analyses were performed using triple quadrupole (QqQ) mass analyzers providing low mass resolution but optimal duty cycle. Product ions of LMs were monitored by using selected reaction monitoring (SRM) and multi reaction monitoring (MRM), respectively. LM identification was based on the specific retention time (RT) of reference compounds and the detection of specific product ions. The development of new MS-analyzer designs like the Orbitrap [18] and HR-TOF MS [19] made accurate mass information compatible with conventional LC as well as UPLC time scales. With these instruments the full MS^2 spectra are recorded and can be utilized for post processing termed as parallel reaction monitoring (PRM) [20]. Hence “mediator lipidomics” [21] or “targeted lipidomics” [22] were introduced for describing the merging of targeted and untargeted approaches. Recently, LC-HR-MS^2 approaches were successfully applied for identification and quantitation of LMs [23,24]. However, metabolite identification based on spectrum-spectrum comparison is a crucial step in all targeted and untargeted approaches [25]. To build up a MS^2 library that provides reliable spectral information of known and unknown compounds, one can benefit from the full HR-MS^2 spectral information obtained by PRM. This has the advantage of decreased levels of false positive lipid identifications, especially when it comes to isomeric LMs [26].

Recently, we developed a SPE-free extraction procedure coupled with a micro-flow LC-ESI-MS^2 [27]. In this study, we provide a comprehensive description of the LC-MS^2 approach using a quadrupole time-of-flight (QqTOF) and a Q Exactice Plus mass spectrometer. We discuss how application of PRM in conjunction with the dedicated software SpeCS can improve quality control of LM quantitation.

2. Materials and methods

2.1. Chemicals

All lipid standards were purchased from Cayman Chemicals (Ann Arbor, Michigan, USA) (see supplementary Table 1). All solvents with at least LC grade were purchased from Sigma-Aldrich (Taufkirchen, Germany) and from Merck Millipore (Darmstadt, Germany). Dulbecco's Modified Eagle Medium (DMEM) and fetal calf serum (FCS) were purchased from Merck Millipore (Darmstadt, Germany). Used antibodies were purchased from BioLegend (San Diego, California, USA). Chemicals for fluorescence and staining assays were obtained from Life Technologies (Carlsbad, California, USA) and Southern Biotech (Birmingham, Alabama, USA), respectively.

2.2. Stock solutions and storages

Reference compounds of oxidized PUFAs and the deuterium-labeled internal standards (IS) for quantitation were dissolved in ethanol, acetonitrile (ACN) or methyl acetate and stored at −20 or −80 °C as recommended by Cayman Chemicals (Ann Arbor, Michigan, USA). All stock solutions had a concentration of 10 pmol/μl. Prostaglandins (PGs) and tri-hydroxylated PUFAs (PUFA-O3), were quantified by using PGE2-d9 and RvD2-d5, respectively. The groups of di- and mono-hydroxylated PUFAs (PUFA-O2, PUFA-O1) were quantified utilizing LTB4-d4 and 5-HETE-d8. The precursor PUFAs were quantitated using AA-d11. For each sample set extraction, new IS mixtures with a concentration of 500 fmol/μl for each deuterated compound were prepared.

2.3. Mycobacteria tuberculosis infection model

Mycobacterium tuberculosis H37Rv (*M. tuberculosis*) was grown in Middlebrook 7H9 broth (BD Biosciences) supplemented with OADC (Oleic acid, Albumin, Dextrose, Catalase) enrichment medium (BD Biosciences). Bacterial cultures were harvested and aliquots were frozen at −80 °C until use. For aerosol infection, *M. tuberculosis* stocks were diluted in sterile distilled water. Animals were exposed to an aerosol containing 200 CFU for 60 min using an aerosol chamber (Glas-Col, Terre-Haute, IN, USA) and bacterial burdens in lung and spleen were evaluated at various time points p.i. as described before [28]. Lung homogenates have been formed by mechanical disruption of the organ in 0.05% v/v Tween 20 in PBS. Lung homogenates have been sterile filtered and were utilized for LM extraction.

2.4. Lipid extraction

Lipids were extracted according to a modified acidified Bligh&Dyer protocol [16,27]. Briefly, 500 μl of the lung homogenates were transferred into an ice-cooled 2.0 ml tube covered with aluminum foil. Afterwards, 2 μL of butyl hydroxyl-toluene in methanol (BHT; 1.0 mg/mL) and 25 μl of the internal standard mix was added to the tube (Table 1). Afterwards, 320 μL of ice-cold chloroform was added followed by 640 μl of ice-cold MeOH/HAc (97:3; v/v) to the mixture. The mixture was incubated for 30 min at room temperature with constant stirring. For phase separation 320 μL of ice-cold water was added and the mixture was incubated for another 30 min under shaking. Finally the samples were centrifuged at 15,000 × g for 15 min at 4 °C. The organic phase was collected and transferred to an ice-cooled sample tube. The extraction procedure was three more times repeated by adding 320 μL of ice-cold $CHCl_3$. Finally, all combined organic phases were dried under a slight stream of nitrogen. Lipid extracts were dissolved in 40 μL of solvent and 2 μL were injected for the LC-MS^2 analysis.

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

Table 1
Retention time (RT), product ions, linear range of reference compounds and limit of detection (LOD), limit of quantification (LOQ) of deuterium-labeled internal standards (IS).

Eicosanoids/Docosanoids	IS	Lipid species	RT (min)	RSD (%)	Precursor ion (*m*/*z*)	Product ion (*m*/*z*)	CE	LOD*	LOQ*	Linear range*	R^2
PUFA-O3	RvD2-d5		2.94	2.4	380.3	175.09	20	0.03	0.11		
PG	PGE2-d9		2.75	2.7	360.3	189.14	17	0.01	0.02		
		RvE1	1.85	4.0	349.2	195.12	17			0.25–25	0.9952
		20-OH-LTB4	1.86	4.0	351.2	195.12	17			0.05–25	0.9968
		LxA5	2.55	5.4	349.2	215.15	17			0.05–25	0.9956
		LxB4	2.70	4.2	351.2	221.12 + 155.08	17			0.05–25	0.9946
		PGE2	2.81	4.3	351.2	175.12 + 207.12	17			0.05–25	0.9985
		RvD2	2.96	4.3	375.2	175.09	20			0.05–25	0.9994
		PGD2	3.04	3.6	351.2	233.13	17			0.05–25	0.9987
		LxA4	3.20	6.1	351.2	115.05 + 135.09	17			0.125–12.5	0.9905
		RvD1	3.40	3.8	375.2	233.17	20			0.25–25	0.9953
		13,14-DiH-15k-PGE2	4.02	3.5	351.2	209.12	17			0.25–25	0.993
		13,14-DiH-15k-PGD2	4.88	2.8	351.2	207.11	17			0.050–25	0.9996
PUFA-O2	LTB4-d4		7.97	3.1	339.3	197.14	16	0.10	0.33		
		5,15-DiHETE	7.31	3.3	335.2	255.23	16			0.05–25	0.9991
		17,18-DiHETE	7.67	4.0	335.2	247.17	16			0.05–25	0.9973
		7-Mar1	7.77	3.5	359.2	250.15 + 177.15	14.5			0.125–25	0.9972
		LTB4	8.12	2.9	335.2	195.12	16			0.05–25	0.9993
		PD1	8.38	3.8	359.2	153.11	14.5			0.05–25	0.9973
		14,15-DiHETE	9.29	2.8	335.2	207.16	16			0.417–25	0.9857
		12-oxo-LTB4	11.52	4.4	333.2	179.0_.2 + 155.07	16			0.25–25	0.9844
PUFA-O1	5-HETE-d8		35.78	0.2	327.2	116.06	17	0.04	0.14		
		18-HEPE	26.25	3.2	317.2	255.23	17			1.25–25	0.9864
		15-HEPE	32.89	2.2	317.2	219.14	17			0.417–25	0.9955
		12-HEPE	34.38	0.4	317.2	179.12	17			0.05–25	0.9989
		17,18-EpETE	35.33	0.1	317.2	215.16	17			0.25–25	0.9956
		15-HETE	35.34	0.2	319.2	175.17	14			0.05–25	0.9965
		17-HDoHE	35.47	0.3	343.2	229.15	13.5			1.25–25	0.9972
		13-HDoHE	35.57	0.2	343.2	193.15	13.5			0.05–25	0.9959
		10-HDoHE	35.63	0.2	343.2	153.11	13.5			0.05–25	0.9964
		14-HDoHE	35.64	0.2	343.2	205.15	13.5			0.05–25	0.994
		8-HETE	35.69	0.2	319.2	155.08	14			0.05–25	0.9985
		12-HETE	35.70	0.2	319.2	179.13	14			0.05–25	0.998
		7-HDoHE	35.74	0.2	343.2	141.07	13.5			0.05–25	0.9967
		5-HETE	35.83	0.2	319.2	115.05	14			0.05–25	0.9973
		4-HDoHE	35.99	0.2	343.2	101.04	13.5			0.125–25	0.9944
		14,15-EET	36.14	0.2	319.2	175.17	14			0.05–25	0.9983
		11,12-EET	36.37	0.2	319.2	167.13	14			0.05–25	0.998
		5,6-EET	36.41	0.2	319.2	191.18	14			0.05–25	0.9973
		8,9-EET	36.42	0.2	319.2	123.08	14			0.05–25	0.9953
PUFA	AA-d11		37.86	0.2	314.3	270.33, 216.29	14	0.03	0.11		
		EPA	37.34	0.2	301.3	257.26, 203.2	11			0.05–25	0.9956
		DHA	37.75	0.3	327.3	283.27, 229.23	14			0.05–25	0.9923
		AA	37.87	0.2	303.2	205.21, 259.27	14			0.05–25	0.9987

* pmol on column.

2.5. LC-MS^2 analysis

Liquid chromatography was performed using a micro-LC 1100 system Agilent Technologies (Santa Clara, California, USA). For chromatographic separation a Luna C18(2) reversed phase (RP) column (0.3 mm ID x 150 mm) from Phenomenex (Torrance, California, USA) was used at a flow rate of 10 µl/min. Solvent A comprised of water, ACN and 1 M ammonium acetate (59:40:1, v/v/v) and the pH was set to 5.6 using acetic acid (HAc). Solvent B was mixed from methanol (MeOH), methyl *tert*-butyl ether (MTBE), 2-propanol and HAc (50:40:10:0.1, v/v/v/v). The LC gradient was set to 0 min: 0% (B); 5 min: 0% (B); 10 min: 100% (B); 30 min: 100%; 35 min: 0% (B); 70 min: 0% (B). Samples were kept at 5 °C in the autosampler and the injection volume was 2.0 µl MS^2 data was recorded on a Q-TOF Ultima™ mass spectrometer from Waters (Milford, Massachusetts, USA) via the Z-spray ESI source. All experiments were performed in the negative ion mode using a capillary voltage of 2.6 kV, cone voltage 65 V and RF lens 1 at 32.5 V. The source and desolvation temperature were set to 100 and 210 °C, respectively. The cone and desolvation gas flow was set 180 and 250 L/h, respectively. Precursor isolation was performed with 1 Da selection window and collision induced dissociation (CID) was achieved using argon as collision gas using adjusted collision energies (CE) for each precursor ion group (Table 1). Full MS^2 spectra were recorded according to a scheduled selection of precursor masses to enable sufficient sampling of chromatographic peaks. Peak detection, integration and quantitation of compounds was performed using MassLynx™ 4.0, Waters (Milford, Massachusetts, USA) allowing mass error of 500 ppm and 1 Da, respectively.

The HR-MS^2 was acquired on a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). All MS experiments were performed in the negative ion mode using a spray voltage of 3.0 kV, a capillary temperature of 260 °C, sheath gas pressure of 5 au and the S-lens radiofrequency level of 100. Precursor isolation was performed with 1 Da selection window and the resolution was set to a nominal resolution of 35,000 (FWHM at *m/z* 200). MS data analysis was performed using Xcalibur software (Thermo Fisher Scientific, Bremen, Germany).

2.6. Determination of linear range

All 39 reference compounds were mixed with the IS mixture at a concentration of 500 fmol/µl. All reference compounds were analyzed in triplicates at concentrations of 10, 25, 50, 85, 100, 500,

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

1000, 1500, 2500 and 5000 fmol/μl. The peak areas for each concentration was utilized for regression analysis to determine the coefficient of determination (R^2) and applicable linear range (Table 1).

2.7. Limit of detection and limit of quantification

To estimate the limit of detection (LOD) and the limit of quantification (LOQ), PGE2-d9, RvD2-d5, LTB4-d4, 5-HETE-d8 and AA-d11 were used as reference compounds for each PUFA group. The standard deviation (SD) of the peak area at the lowest detectable concentration was used to model the noise of the MCP-detector. For the estimation of LOD and LOQ the slope of the calibration curve was multiplied with 3xSD and 10xSD, respectively. All samples were analyzed in triplicates. The same workflow was used for extracted DMEM (250 μl) containing 10% FCS later on.

2.8. Determination of matrix effect and extraction efficiency

Matrix effect and extraction efficiency of the IS mixture (amount: 12.5 pmol) were determined in 250 μl of DMEM containing 10% FCS as described previously [15]. Sample preparation was done as described above. The matrix effect was calculated as the peak areas based ratio in percentage of the spiked IS post-extraction and the pure IS. The extraction efficiency was determined as the peak areas based ratio in percentage of the spiked IS pre- and post-extraction.

2.9. Tandem-MS analysis of 12-HETE and 11,12 EET

Negative mode Electrospray-(ESI)-MS^2 experiments were conducted on an LTQ Orbitrap hybrid instrument [29]. Lipid mediators were diluted in methanol (250 fmoL/μL concentrations) and introduced into the ion source via a syringe pump (flow rate 5 μL/min). Spray voltages were typically 2.5−3.5 kV. The ESI heater temperature was set to 50 °C and the capillary temperature to 275 °C. To generate stable spray conditions, sheath and sweep gases were used (≥99.999% N_2). Monoisotopic precursor ion selection was performed in the LTQ for all product ion CID experiments. MS^2-product ion experiments (collision gas He ≥ 99.999%) were performed in the LTQ with a collision energy adjusted to achieve extensive fragmentations [30,31]. Exact masses of the precursor and product ions were measured in the Orbitrap analyzer (R = 30.000, FWHM @ *m/z* 400), which was externally calibrated with caffeine, trileucine, and thymopentin. All of the presented ion structures are consistent with the experimentally determined exact ion masses (Δm ≤ 2 ppm) and match the theoretical isotope distributions. Data acquisition was conducted with the Tune Plus and data processing and evaluation with Qualbrowser.

2.10. Spectral comparison score (SCS) application for lipid mediators

Each of the 36 reference compounds, excluding the three PUFAs and the five deuterium-labeled IS, were dissolved at a concentration of 250 fmol/μl in the solvent composition according to its retention time and applied LC-gradient. All compounds were analyzed in direct infusion (DI) at a flow rate of 10 μl/min. The averaged MS^2 spectrum was computed from 0.5 min acquisition time and three technical replicates were recorded. Peak lists were analyzed using the software package SpeCS (Available at https://lifs.isas.de/tools/specs). After excluding the precursor and all *m/z* values not matching the fragment ion sum composition constraint $C_xH_yO_{0\text{-}5}$ (positive list), spectrum correlation score (SCS) were calculated according to equation (1). The SCS from forward and reverse spectrum-spectrum matches (SSMs) were averaged for Fig. 3.

Pairs of AA derived isomers, 8-HETE/8,9-EET, 12-HETE/11,12-EET and 15-HETE/14,15-EET were mixed in different ratios at an overall concentration of 1 pmol/μl. Mixed samples were analyzed with DI-ESI-QqTOF. Afterwards, SCS were computed using full MS^2 spectra.

3. Results and discussion

3.1. Lipid extraction and characterization of the LC-MS^2 approach

We established an LC-MS^2 approach that can have a simpler sample preparation procedure avoiding the time consuming SPE enrichment step. For this purpose, an earlier described LC gradient based on an acidified water/acetonitrile [14,16] was modified by adding MTBE, a strong lipophilic solvent, and 2-propanol. With this solvent composition, we could largely remove remaining membrane lipids as well as neutral lipids from the column [15,26,32]. We showed recently that this strategy is applicable for complex matrices like skin biopsies [27]. Here, we describe in detail the chromatographic behavior using 44 reference compounds of eico- and docosanoids (Fig. 1). We observed that the degree of oxidation, accounted in the number of additional oxygen, results in four distinguishable retention time windows: PUFA (36−40min), PUFA-O1 (25−36 min), PUFA-O2 (RT: 7−15 min), PUFA-O3 including prostaglandins (PG) (RT: 2−6 min). For quantitation, five deuterated standards were chosen that belong to one of the four RT windows, namely: PGE2-d9, RvD2-d5, LTB4-d4, 5-HETE-d8 and AA-d11. LODs of these deuterated standards were determined in the range of 10−100 fmol on column (Table 1). Afterwards, each RT specific standard compound was utilized to determine the applicable quantitation range of all LMs and PUFAs with a specific focus to utilize only fragment ions that are of highest specificity. In this regard, we determined an acceptable linear range for most of the molecules of around 50 fmol on column and $R^2 \geq 0.98$ for 39 compounds of interest (Table 1).

Next, we tested the performance of the LC-MS^2 method in combination with a SPE-free extraction procedure by adding known amounts of the deuterated standards to 250 μl of DMEM performing the complete workflow. Experiments were conducted as described previously [15]. We could achieve matrix effects of 20−30% for PGE2-d9, RvD2-d5 and LTB4-d4, which is 15−35% better than reported for colon matrix. Extraction yields were determined to be approx. 100% for all IS (Supplementary Table 2). From these data we concluded that the SPE-free workflow and LC gradient are suitable approaches for LM analysis and are comparable to LC-QqQ based methods published previously [14,16].

We utilized a QqTOF system to perform PRM that record all fragment ions in the range of *m/z* 100−400. All MS spectra and peak lists of the analyzed reference compounds are provided as resource (Supplementary Data 1). Such measurements enables us, to emulate different mass resolution and accuracy for selected transitions utilized for quantification. When we analyzed all 44 standards, with mass accuracy of 1 Da overlapping quantifier ions within the standard compound mixture were detectable (Supplementary Figure 1). Most prominently, one could recognize unresolved pairs of isomeric AA derivatives, 15-HETE/14,15-EET (*m/z* 319.2 → 175.1), 12-HETE/11,12-EET (*m/z* 319.2 → 179.1) and 8-HETE/8,9-EET (*m/z* 319.2 → 155.1). However, if we took advantage of the specificity of the TOF − analyzer, applying an accuracy of 500 ppm such overlaps could be clearly minimized as shown in Fig. 1. A closer inspection of the extracted ion chromatograms and recorded MS^2 spectra points to an additional complexity that the choice of specific fragment ions for quantitation (quantifier ions) has to be performed with sophistication. In a biological samples

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

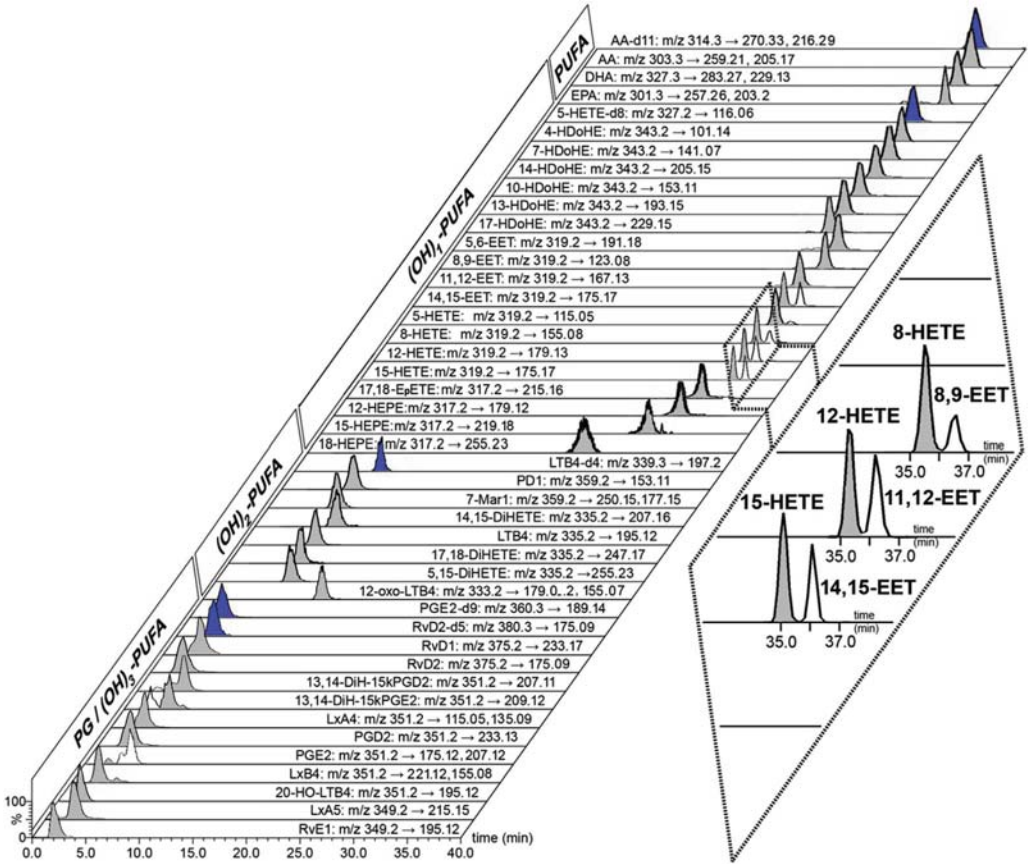


Fig. 1. Representative extracted ion chromatograms of 44 lipid mediators. Negative ESI-Q-TOF-MS^2 were acquired in the mass range *m/z* 100–400 for all indicated precursor ions. All extracted ion chromatograms (XICs) for selected quantifier ions were determined with mass accuracy of 500 ppm. PUFAs and the LM groups PUFA-O1, PUFA-O2, PUFA-O3 and PG have distinct retention time ranges. XICs of the isomeric HETE and EET derivatives are highlighted that have a spectral overlap.

with additional matrix effects one might not be able to independently quantify isomeric LMs because of the chromatographic and mass spectral overlaps. For instance, one can already observe a spectral overlap between PGE2, 13,14-DiH-15kPGD2 and 13,14-DiH-15kPGE2 for the fragment ion *m/z* 175.12, which is present in all three MS^2 spectra and recognizable in the extracted ion chromatograms (Fig. 1, Supplementary Data 1).

3.2. Tandem mass spectrometric analysis and design of spectral library

To address such spectral overlap, which might increasingly occur in complex biological samples, we investigated how the information of the full MS^2 spectra could increase the specificity for LM identification. First, we recorded and compared the DI-MS^2 spectra of 36 lipid mediators (Supplementary Data 1). Exemplary, we discuss the HETE and EET isomers (Fig. 2A). One can recognize that 15-HETE and 14,15-EET spectra are almost identical while 8-HETE/8,9-EET and 12-HETE/11,12-EET exhibit recognizable differences. Although, intensive product ions were selected as quantifier, those were not always decisive for the specificity of MS^2 spectra. In this regard, the quantifier ions commonly used [14,15] for 8-HETE, *m/z* 155.1 were also present in 8,9-EET and for 12-HETE, *m/z* 179.1 was present in 11,12-EET as wells as in 8,9-EET (Fig. 2A). Just from this example, one can acknowledge that a quantification for a set of structurally related LMs can be distorted by spectral and chromatographic overlaps.

Accordingly, we developed the software SpeCS to build spectral libraries, perform spectrum comparisons and to provide a quality control measure for quantification. With SpeCS, peak lists can be processed by applying several filters for intensity threshold, mass accuracy, background signals and residual precursor signals (Supplementary Figure 2A). During the generation of the spectral library we recognized in a number of LMs even numbered fragment ions, for instance for 12-HETE and 11,12EET at *m/z* 208.1 (Fig. 2) or *m/z* 234.126 for HDoHE isomers (Fig. 4A). That was a somehow unexpected result, because only even-electron fragment anions were at first anticipated from CID-MS^2 with a sum composition constraint of $C_xH_yO_{0-5}$. Accordingly, we investigated the fragmentation of 12-HETE and 11,12EET in detail using Q Exactive Plus, Q-TOF, LTQ Orbitrap XL and LTQ XL Iontrap (Supplementary Figure 3). A fragmentation pathway to form radical anion was reported for 8,15-diHETE and 12-HETE by Wheelan et al. following a Oxy-Cope rearrangement [33,34]. Accordingly, we investigated how intensity ratio between *m/z* 208.1 and 207.1 signifying energetically very different fragmentation pathway is affected by the CID characteristics of the utilized mass spectrometer. In this respect, we observed for the Q-TOF, the highest signal intensity for the odd-electron fragment anion *m/z* 208.1 (Supplementary Table 3, Supplementary Figure 3). In contrast, ion trap CID-MS^2 showed much less signal intensity for odd-electron fragment anion. Further

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

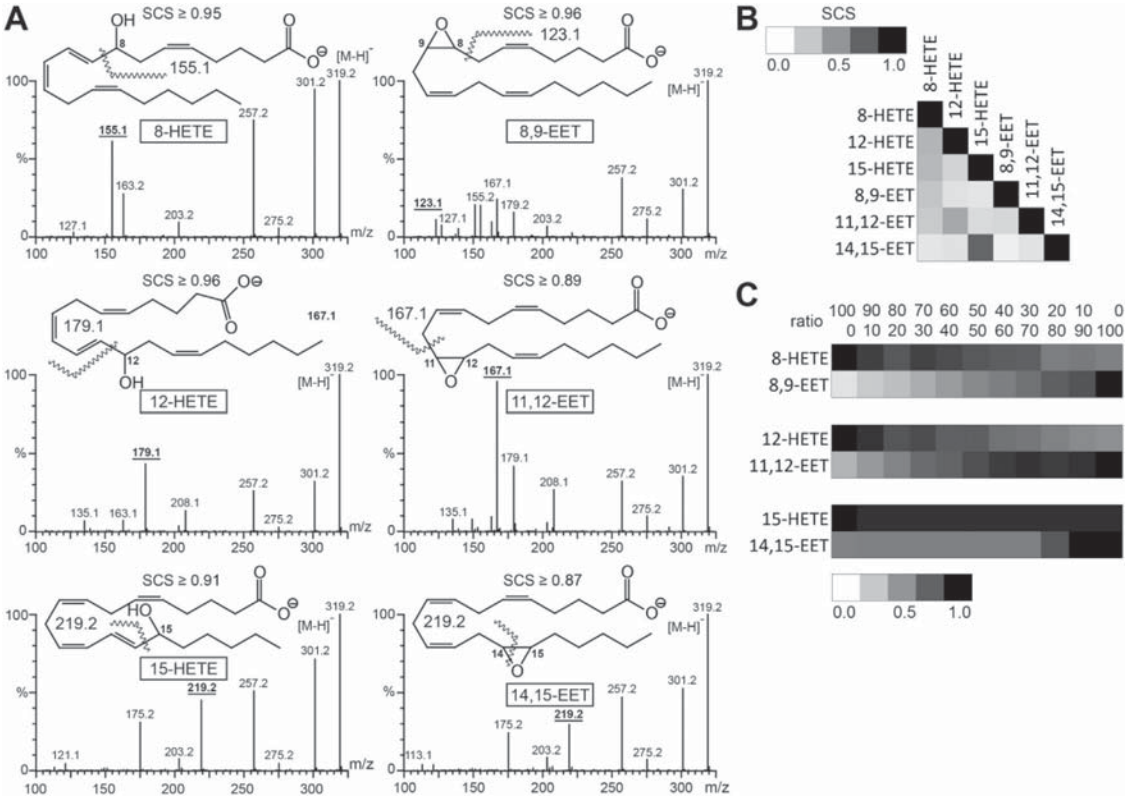


Fig. 2. Spectral comparison scores (SCS) for recognition of isomeric oxidized PUFAs. (A) Full negative ESI-MS^2 spectra (*m/z* 100–400) of the indicated AA-derivatives were shown at a concentration of 250 fmol/µl ([M-H]$^-$: *m/z* 319.2). *m/z* values of applied quantifier ions were underlined and indicated in bold. Typical product ions were [M-H-H_2O]$^-$: 301.2; [M-H-CO_2]$^-$: 275.2; [M-H-CO_2-H_2O]$^-$: 257.2; [M-H-$C_5H_8O_3$]$^-$: 203.2. The lowest spectral scores for three technical replicates is shown on top of each spectra. (B) Heat map of averaged SCS of forward and backward spectrum comparisons. (C) The indicated pairs of isomers were mixed at the given ratio. The obtained spectral data were utilized for forward spectra comparison of the given isomer shown in the respective panel bar.

analysis of all recorded MS^2 spectra showed that structural motifs like 1,2-hydroxy-3-diene and/or 3-hydroxy-1,5-diene in close proximity to conjugated double bonds led to odd-electron fragment ions (Supplementary Figure 3, Supplementary Data 1). Eleven out of 36 recorded lipid mediators comprised odd-electron fragment ions. Accordingly, we added odd-electron fragment ions to the positive filter, which contains all *m/z* values of all possible $C_xH_yO_{0\text{-}5}$ compositions of CID induced fragment ions.

$$SCS = \frac{n_f}{N} r_s^2 \quad (1)$$

nf: number of matched peaks, N: total number of peaks in query, rs Spearman's rank correlation. The implemented algorithm of the SpeCS software is shown in Supplementary Figure 1B.

3.3. Design and characteristics of the spectrum comparison algorithm

The algorithm for computing spectral comparison score (SCS) is based on signal matching as well as comparing signal intensities using Spearman rank correlation (equation (1), Supplementary Figure 1B). In this way the SCS also included a metric for the overall intensity profile of matched signals that gives more weight to intense signals. The SCS, has a well-defined range, where SCS = 0 indicates no match and a SCS = 1 indicates a perfect match. As first test, we evaluated how stable the fragmentation profile for CID-MS^2 are on the Q-TOF mass spectrometer. For three independent measurements, we determined an almost perfect match a spectral comparison score SCS≥0.87 for all molecules (Fig. 2A), which indicated that stable fragmentation pattern were obtained on the Q-TOF instrument platform. For the six exemplary chosen, all pairwise comparison were performed using the SCS algorithm, which was in agreement with the overall similarity between spectra, with the highest score for the comparison between 15-HETE and 14,15-EET with SCS = 0.63 (Fig. 2B).

Next, we explored with DI experiments, how co-elution of such highly similar molecules would affect the scoring algorithm. For that, we acquired MS^2 spectra for each pair of isomers in different ratios (Fig. 2C). The forward SSMs of 8-HETE and 12-HETE with its corresponding EET isomers returned a significantly decreased SCS for the ratio 90:10 of 0.89 and 0.29, respectively (Fig. 2C). The corresponding forward SSMs for 8,9-EET and 11,12-EET (ratio 10:90) resulted in relatively small SCS values of 0.24 and 0.486. Regarding the SSM of 15-HETE with 14,15-EET, one can recognize that the specificity of the scoring mechanism was insufficient, because all major abundant ions were present in both mediators and had very similar intensity profiles (Fig. 2B). However, we concluded from this result that the score could not be utilized to exclude the presence of a structurally close molecule but might help to recognize major contribution of specific LM as quality criteria. Co-elution of LM isomers and the resulting mixed MS^2 spectra was nearly impossible to avoid, when expanding the scope of LM analysis [26].

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

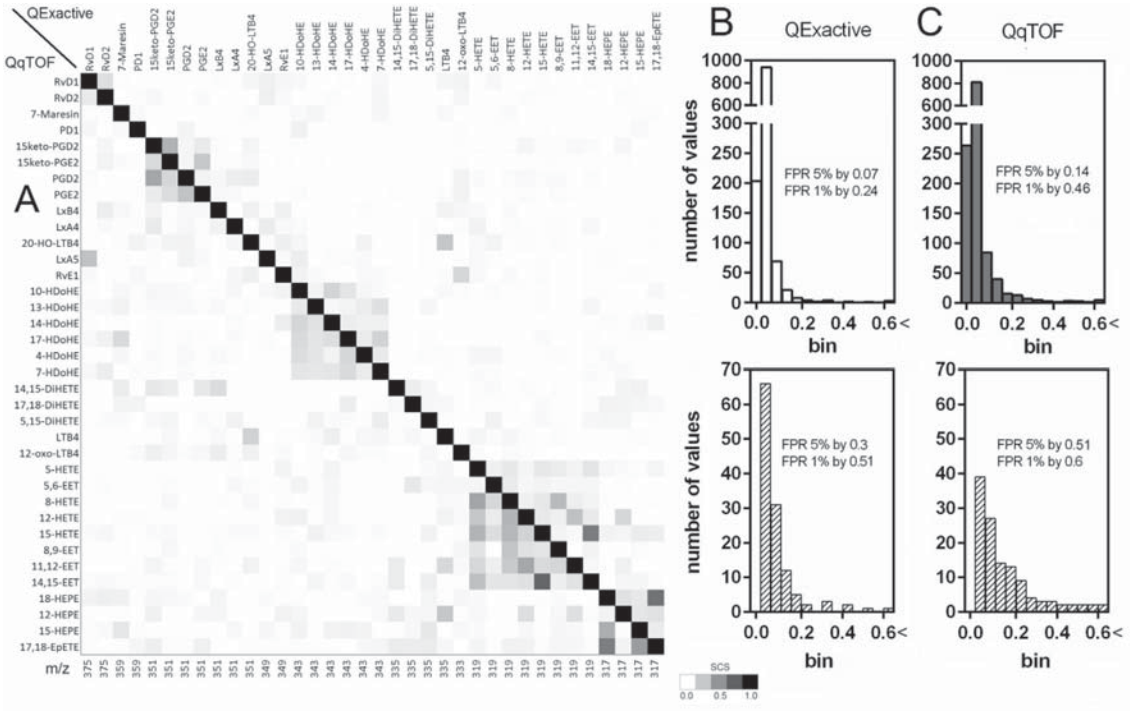


Fig. 3. SCS of oxidized PUFAs as estimator of False Discovery Rates. (A) Heat map for the averaged SCS of 36 reference compounds sorted according to their precursor *m/z* value for the Q-TOF and Q Exactive data. SCS based on the Intensity threshold of 1% and a bin range of 0.1 for Q-TOF data and 0.005 for Q Exactive data were calculated for all precursor ions, omitting different retention on the LC-column. (B), (C) Histogram of all 1260 SCS values and histogram of 128 SCS values only for spectral comparison within each precursor ion group of *m/z* 319, 343 and 351 for the Q Exactive and Q-TOF data respectively.

To further investigate the impact of HR-MS^2 to improve specificity of the LM quantitation using PRM, we compared MS^2 of three isomeric LMs 10-HDoHE, 17-HDoHE and 7-HDoHE on the Q Exactive (Supplementary Figure 4). In particular, fragment ions in the *m/z* region of 121.04–121.14, 189.1–189.2 and 245.12–245.22 allowed a clear differentiation between these isomers, which could not be resolved with the Q-TOF. As result, the specificity of SCS between the six HDoHE isomers was clearly improved using HR-MS^2 reducing the average SCS of all pairwise comparisons from 0.124 (Q-TOF) to 0.081 (Q Exactive) (Fig. 3A). The improved specificity of the SCS was a result of the fact that a mass difference of 0.036 Da is distinguishable, which is characteristic for exchange of an oxygen with CH_4. In this way, one can unequivocally assign sum composition for all possible fragment ions within constraint of $C_xH_yO_{0\text{-}5}$. Therefore, a better differentiation between positional isomers for hydroxyl groups was achieved. We further concluded that resolution of 35,000 (FWHM @ *m/z* 200) and accuracy of 0.005 Da should be sufficient to achieve best scoring specificity.

3.4. Estimation of false discovery rates for PRM based quantitation

To gain insight into the overall performance of the scoring approach in dependence to mass accuracy and resolution we calculated the SCS of 36 reference compounds recorded on the Q-TOF and Q Exactive, respectively. For that, a MS^2 spectrum library was compiled which was made available as a resource together with the SpeCS software package. Subsequently, the SCS for 1260 forward and backward spectrum comparisons was computed using both instrument dependent spectral libraries (Fig. 3). We included all MS^2 spectra even when LMs comprised different precursors *m/z* and had different RTs (Fig. 3A). This strategy was motivated by the fact that the LM share many structural features and the data basis for modeling false positive rates (FPRs) could be extended. In support of this concept, we recognized several LMs with increased SCS like LxA5 (precursor ion: *m/z* 349.2) and RvD1 (precursor ion: *m/z* 375.2) with SCS = 0.523 (Q-TOF), SCS = 0.324 (Q Exactive) as well as LTB4 (precursor ion: *m/z* 335.2) and 20-HO-LTB4 (precursor ion: *m/z* 351.2) with SCS = 0.253 (Q-TOF), SCS = 0.081 (Q Exactive). At the same time, overall higher SCS values were computed for LMs of the same precursors like *m/z* 351.2, 343.2 and 319.2 (Fig. 3A). The distribution of all pairwise SCS values was utilized to estimate false positive rates (FPRs). For a FPR of 5%, a SCS ≥0.14 (Q-TOF) and SCS = 0.07 (Q Exactive) was determined for all SSMs. When we increased the stringency and performed an analysis for 128 SSMs of LMs with precursor *m/z* 351.2, 343.2 and 319.2 a SCS ≥0.51 (Q-TOF) and SCS ≥ 0.6 (Q Exactive) would be sufficient to reach a FPR of 1% (Fig. 3 B, C).

From these results, we concluded that the SCS could be utilized as quality criteria to which degree a quantitation, based on a single specific fragment ion, originated from one LM alone. We would approximate that SCS ≥0.24 (Q-TOF), SCS ≥ 0.46 (Q Exactive) for FPR: 1% indicates the presence of a structurally related LM during quantitation. When a score of SCS ≥0.8 was calculated the quantitation was mostly likely based upon a single analyte.

3.5. Quality control of HDoHE isomers quantitation during M. tuberculosis infection

To investigate the behavior of the SCS algorithm in a biological application, we profiled DHA-derived oxidized PUFAs in the course of *M. tuberculosis* infection. In this case we recorded the full MS^2 spectra

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

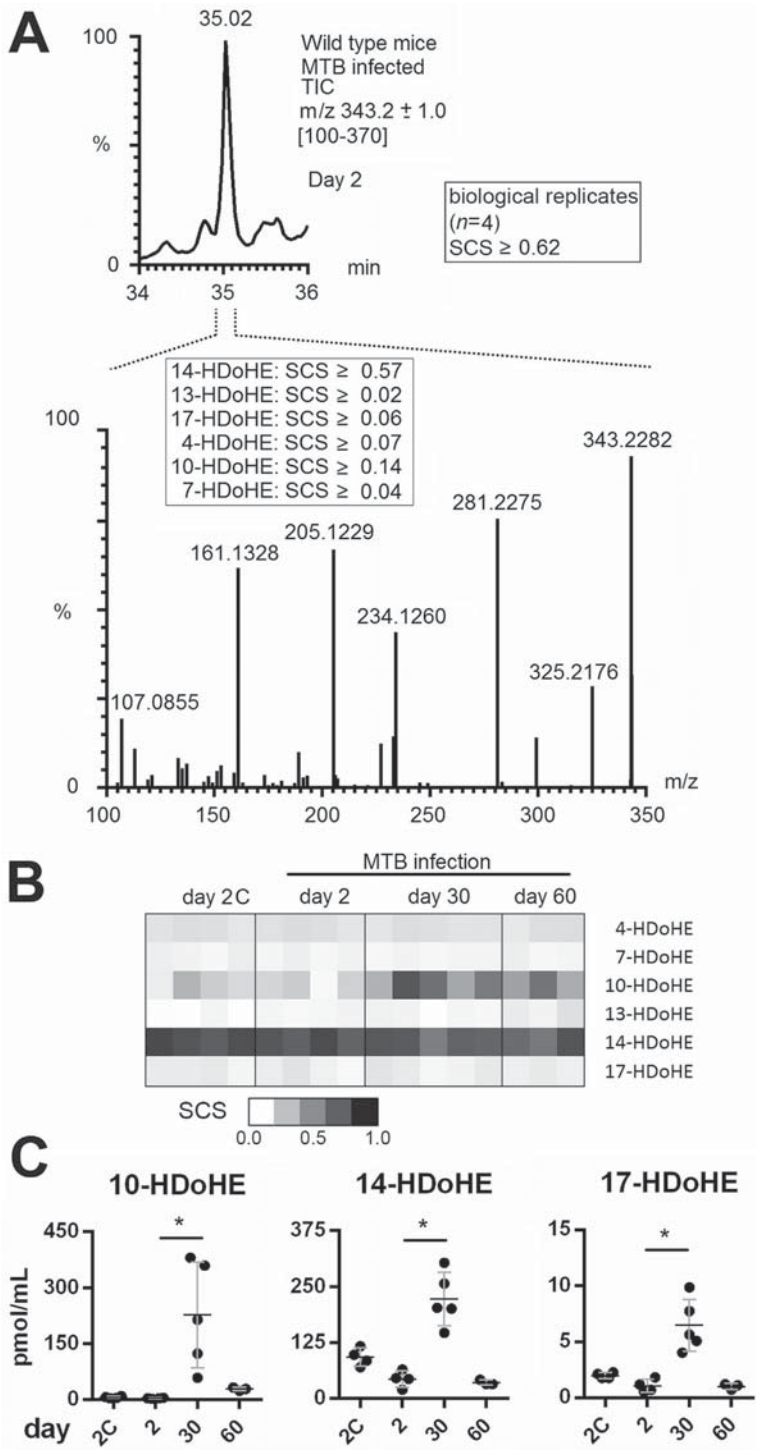


Fig. 4. Spectral comparison analysis of HDoHE isomers during *M. tuberculosis*. (A) Averaged ESI-MS^2 spectra of mono-hydroxylated DHA-derivatives obtained at the RT range of the precursor peak *m/z*: 343.2 acquired on the Q Exactive respectively (B) SCS were calculated for each biological replicate during time course of infection and were summarized in the heat map. (C) Absolute quantitation of 10-HDoHE, 14-HDoHE and 17-HDoHE using the quantifier *m/z* 153.1, *m/z* 193.1 and *m/z* 205.1, respectively. ($n = 4$–5; stars indicate significance in the two-tailed Mann-Whitney *U* Test with *, $p < 0.05$, **, $p < 0.005$).

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044

on the Q Exactive and we were able to quantify 10-HDoHE, 13-HDoHE, 14-HDoHE and 17-HDoHE using PRM (Fig. 4A). For these molecules a clear LC-separation was impossible and RT difference below 12 s were detected. When we determined the SCS between the averaged MS^2 spectra of biological replicates we retrieved scores of SCS $\geq$0.62 (Fig. 4A). By applying our score over all time points we observed that 14-HDoHE was the most abundant LM with a score in a range of 0.43 until 0.69. The scores indicated a high probability that similar molecules were present which was in conjunction with the solely quantifier based result. We further noticed that the SCS for 10-HDoHE showed increased concentrations from day 30 onwards while all other derivate showed no significant changes (Fig. 4B). The changes in the score for 10-HDoHE were in good agreement with the quantification results. As a negative control we computed SCS for 4-HDoHE, which was below the LOQ if only the quantifier ion *m/z* 101.036 was considered. Accordingly, we had on average over all time points a SCS = 0.07. In case of 17-HDoHE the quantification results showed significant increased concentration at day 30, which was not reflected in the score. However, concentration of 17-HDoHE at day 30 was still approximately 10 times lower than 10-HDoHE and 14-HDoHE that had only negligible influence on the profile of the mixed MS^2 spectrum.

4. Conclusion

We described a SPE-free, LC-MS^2 based method for the quantitation of LMs and PUFAs. The combination of PRM quantitation with the SCS approach enabled to introduce a well-defined quality control for LM quantitation. In concert with incorporated filters for peak list processing, we evaluated specifically how structurally related molecules and isomeric LMs could be recognized, which are difficult to separate by LC. In this study we specifically focused on the analytical separation power of the mass spectrometer. Certain mass spectrometric overlaps can be addressed by improving the LC-conditions. However, we specifically wanted to investigate how much HR-MS^2 can help to improve the quality of lipid mediator quantification. Our approach is adjustable to different instrument platforms and LC method and can be successively expanded in its scope by increasing the size of the underlying spectral library. The SpeCS software was designed in such a way that spectrum libraries for other compound classes could be implemented. We anticipate that this feature might support the community in building dedicated spectral libraries of metabolites and will help to design and exchange quantitation approaches based on PRM and LC-HR-MS^2.

Acknowledgements

This work was supported by funds of the German Research Foundation (Cluster of Excellence 306, CL X) and the Federal Ministry of Education and Research (German Network for Bioinformatics Infrastructure (de.NBI), Lipidomics Informatics for Life Sciences (LIFS, grant number: 031L0108B)).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.aca.2018.01.044.

References

[1] R.J. Helliwell, L.F. Adams, M.D. Mitchell, Prostaglandin synthases: recent developments and a novel hypothesis, Prostaglandins Leukot. Essent. Fatty Acids 70 (2) (2004) 101–113.
[2] H. Kuhn, V.B. O'Donnell, Inflammation and immune regulation by 12/15-lipoxygenases, Prog. Lipid Res. 45 (4) (2006) 334–356.
[3] W.L. Smith, Nutritionally essential fatty acids and biologically indispensable cyclooxygenases, Trends Biochem. Sci. 33 (1) (2008) 27–37.
[4] C.N. Serhan, N. Chiang, T.E. Van Dyke, Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators, Nat. Rev. Immunol. 8 (5) (2008) 349–361.
[5] C.D. Buckley, D.W. Gilroy, C.N. Serhan, Proresolving lipid mediators and mechanisms in the resolution of acute inflammation, Immunity 40 (3) (2014) 315–327.
[6] C.N. Serhan, et al., Protectins and maresins: new pro-resolving families of mediators in acute inflammation and resolution bioactive metabolome, Biochim. Biophys. Acta 1851 (4) (2015) 397–413.
[7] C.D. Sadik, T. Sezin, N.D. Kim, Leukotrienes orchestrating allergic skin inflammation, Exp. Dermatol. 22 (11) (2013) 705–709.
[8] T. Sezin, D. Zillikens, C.D. Sadik, Leukotrienes do not modulate the course of Aldara-induced psoriasiform dermatitis in mice, Acta Derm. Venereol. 95 (3) (2015) 341–342.
[9] C.D. Sadik, A.D. Luster, Lipid-cytokine-chemokine cascades orchestrate leukocyte recruitment in inflammation, J. Leukoc. Biol. 91 (2) (2012) 207–215.
[10] A.C. Kendall, A. Nicolaou, Bioactive lipid mediators in skin inflammation and immunity, Prog. Lipid Res. 52 (1) (2013) 141–164.
[11] V.C. Tam, et al., Lipidomic profiling of influenza infection identifies mediators that induce and resolve inflammation, Cell 154 (1) (2013) 213–227.
[12] J.S. Dickinson, R.C. Murphy, Mass spectrometric analysis of leukotriene A4 and other chemically reactive metabolites of arachidonic acid, J. Am. Soc. Mass Spectrom. 13 (10) (2002) 1227–1234.
[13] R. Deems, et al., Detection and quantitation of eicosanoids via high performance liquid chromatography-electrospray ionization-mass spectrometry, Methods Enzymol. 432 (2007) 59–82.
[14] J. Yang, et al., Quantitative profiling method for oxylipin metabolome by liquid chromatography electrospray ionization tandem mass spectrometry, Anal. Chem. 81 (19) (2009) 8085–8093.
[15] P. Le Faouder, et al., LC-MS/MS method for rapid and concomitant quantification of pro-inflammatory and pro-resolving polyunsaturated fatty acid metabolites, J Chromatogr B Analyt Technol Biomed Life Sci 932 (2013) 123–133.
[16] X. Liu, et al., Oxidized fatty acid analysis by charge-switch derivatization, selected reaction monitoring, and accurate mass quantitation, Anal. Biochem. 442 (1) (2013) 40–50.
[17] A. Sasaki, et al., Determination of omega-6 and omega-3 PUFA metabolites in human urine samples using UPLC/MS/MS, Anal. Bioanal. Chem. 407 (6) (2015) 1625–1639.
[18] A. Makarov, Electrostatic axially harmonic orbital Trapping: a high-performance technique of mass analysis, Anal. Chem. 72 (6) (2000) 1156–1162.
[19] P.F. DeCarlo, et al., Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer, Anal. Chem. 78 (24) (2006) 8281–8289.
[20] S. Gallien, et al., Selectivity of LC-MS/MS analysis: implication for proteomics experiments, J Proteomics 81 (2013) 148–158.
[21] K.A. Massey, A. Nicolaou, Lipidomics of oxidized polyunsaturated fatty acids, Free Radic. Biol. Med. 59 (2013) 45–55.
[22] G. Astarita, et al., Targeted lipidomic strategies for oxygenated metabolites of polyunsaturated fatty acids, Biochim. Biophys. Acta 1851 (4) (2015) 456–468.
[23] T. Sanaki, et al., A hybrid strategy using global analysis of oxidized fatty acids and bioconversion by *Bacillus circulans*, Rapid Commun. Mass Spectrom. 30 (6) (2016) 751–762.
[24] M. Masoodi, et al., Comprehensive lipidomics analysis of bioactive lipids in complex regulatory networks, Anal. Chem. 82 (19) (2010) 8176–8185.
[25] Y. Ma, et al., MS2Analyzer: a software for small molecule substructure annotations from accurate tandem mass spectra, Anal. Chem. 86 (21) (2014) 10724–10731.
[26] G. Dasilva, et al., Lipidomic analysis of polyunsaturated fatty acids and their oxygenated metabolites in plasma by solid-phase extraction followed by LC-MS, Anal. Bioanal. Chem. 406 (12) (2014) 2827–2839.
[27] T. Sezin, et al., The leukotriene B4 and its receptor BLT1 act as critical drivers of neutrophil recruitment in murine Bullous Pemphigoid-like Epidermolysis Bullosa acquisita, J. Invest. Dermatol. 137 (5) (2017) 1104–1113.
[28] B.E. Schneider, et al., Lysosomal phospholipase A2: a novel player in host immunity to *Mycobacterium tuberculosis*, Eur. J. Immunol. 44 (8) (2014) 2394–2404.
[29] A. Makarov, et al., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, J. Am. Soc. Mass Spectrom. 17 (7) (2006) 977–982.
[30] O.V. Borisov, et al., Low-energy collision-induced dissociation fragmentation analysis of cysteinyl-modified peptides, Anal. Chem. 74 (10) (2002) 2284–2292.
[31] L.L. Lopez, et al., Automated strategies for obtaining standardized collisionally induced dissociation spectra on a benchtop ion trap mass spectrometer, Rapid Commun. Mass Spectrom. 13 (8) (1999) 663–668.
[32] S. Hebbar, et al., Laser capture microdissection coupled with on-column extraction LC-MS(n) enables lipidomics of fluorescently labeled Drosophila neurons, Anal. Chem. 86 (11) (2014) 5345–5352.
[33] P. Wheelan, J.A. Zirrolli, R.C. Murphy, Low-energy fast atom bombardment tandem mass spectrometry of monohydroxy substituted unsaturated fatty acids, Biol. Mass Spectrom. 22 (8) (1993) 465–473.
[34] P. Wheelan, J.A. Zirrolli, R.C. Murphy, Electrospray ionization and low energy tandem mass spectrometry of polyhydroxy unsaturated fatty acids, J. Am. Soc. Mass Spectrom. 7 (2) (1996) 140–149.

Please cite this article in press as: A. Wutkowski, et al., Software-aided quality control of parallel reaction monitoring based quantitation of lipid mediators, Analytica Chimica Acta (2018), https://doi.org/10.1016/j.aca.2018.01.044